Please type a plus (+) inside this box → ⊞

PTO/SB/29 (12/97)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.	7326-101
Total Pages	99
First Named Inventor or Application Identifier	Costa, Michael R.
Express Mail Label No.	EM 061 021 415 US

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO: **Assistant Commissioner for Patents**
Box Patent Application
Washington, DC 20231

1. ☒ Fee Transmittal Form (in duplicate) [*Total Pages* 2]
 Submit an original, and a duplicate for fee processing)
2. ☒ Specification (including Table of Contents) [*Total Pages* 70]
 (preferred arrangement set forth below)
 - -Descriptive title of the Invention
 - -Cross Reference to Related Applications
 - -Statement Regarding Fed sponsored R&D
 - -Reference to Microfiche Appendix
 - -Background of the Invention
 - -Brief Summary of the Invention
 - -Brief Description of the Drawings *(if filed)*
 - -Detailed Description of the Invention (including drawings, *if filed*)
 - -Claim(s)
 - -Abstract of the Disclosure
3. ☒ Drawing(s) *(35 USC 113)* (10 Figures) [*Total Sheets* 25]
4. ☒ Oath or Declaration (unexecuted) [*Total Sheets* 2]
 a. ☐ Newly executed (original or copy)
 b. ☐ Copy from a prior application (37 CFR 1.63(d))
 (for continuation/divisional with Box 17 completed)
 [Note Box 5 below]
 i. ☐ DELETION OF INVENTORS(S)
 Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33 (b).
5. ☐ Incorporation By Reference *(useable if Box 4b is checked)*
 The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program *(Appendix)*
7. ☐ Nucleotide and/or Amino Acid Sequence Submission
 (if applicable, all necessary) [Total Pages]
 a. ☐ Computer Readable Copy
 b. ☐ Paper Copy (identical to computer copy)
 c. ☐ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney
 (when there is an assignee)
10. ☐ English Translation Document *(if applicable)*
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)
 (Should be specifically itemized)
14. ☐ Small Entity ☐ Statement filed in prior application, Statement(s) Status still proper and desired
15. ☐ Certified Copy of Priority Document(s)
 (if foreign priority is claimed)
16. ☐ Other:

17. **If a CONTINUING APPLICATION,** *check appropriate box and supply the requisite information:*
☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: ϕϕ filed ϕϕ.

18. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label 20583 *(Insert Customer No. or Attach bar code label here)* or ☐ Correspondence address below

NAME					
ADDRESS					
CITY		*STATE*		*ZIP CODE*	
COUNTRY		*TELEPHONE*		*FAX*	

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

Attorney Docket No. 7326-101

Express Mail No. EM 061 021 415 US

ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

TABLE OF CONTENTS

FIELD OF THE INVENTION 1

BACKGROUND OF THE INVENTION 1

SUMMARY OF THE INVENTION 4

BRIEF DESCRIPTION OF THE DRAWINGS 5

DETAILED DESCRIPTION OF THE INVENTION 5

EXAMPLE 1: CLONING OF *C. ELEGANS* SREBP 36

EXAMPLE 2: ceSREBP EXPRESSION ANALYSIS 42

EXAMPLE 3: RNA INTERFERENCE (RNAI) OF *C. ELEGANS* SREBP, S2P and SCAP 44

EXAMPLE 4: DOMINANT NEGATIVE ceSREBP PHENOTYPES 49

EXAMPLE 5: TC1 TRANSPOSON MUTAGENESIS 50

EXAMPLE 6: CLONING OF *DROSOPHILA* S2P 56

EXAMPLE 7: CLONING OF *DROSOPHILA* SCAP 57

EXAMPLE 8: TRANSGENIC DROSOPHILA MISEXPRESSING SREBP 59

ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

FIELD OF THE INVENTION

The present invention relates to animal models useful for the study of lipid metabolism that have been genetically modified to express or mis-express proteins involved in the sterol regulatory element binding protein (SREBP) pathway. The invention also relates to novel SREBP pathway nucleic acid and polypeptide sequences and their uses.

BACKGROUND OF THE INVENTION

Triglycerides, phospholipids, and cholesterol, which form the three major classes of lipid, perform a variety of necessary functions in cell metabolism and are vital constituents of biological membranes. However, elevated levels of lipids and/or improper lipid metabolism have been implicated in a variety of health disorders. Of particular concern is increased blood cholesterol which leads to atherosclerosis (the deposition of cholesterol on arterial walls). This in turn may lead to heart disease, stroke or other disorders of the circulatory system. Accordingly, there is much interest within the pharmaceutical industry to understand the mechanisms involved in cholesterol synthesis and metabolism, particularly on the molecular level, so that blood cholesterol lowering drugs can be developed for the treatment or prevention of atherosclerosis.

Recent advances have been made in understanding some of the mechanisms involved in mammalian lipid metabolism. A key component is the sterol regulatory element binding protein (SREBP) pathway. SREBPs are transcription factors that activate genes involved in cholesterol and fatty acid metabolism. In the cholesterol biosynthetic pathway of vertebrates, SREBPs directly activate transcription of the genes encoding 3-hydroxy-3-methylglutaryl (HMG) coenzyme A synthase, HMG-CoA reductase, farnesyl diphosphate synthase, and squalene synthase. In the fatty acid and triglyceride biosynthetic pathways, the direct targets of SREBPs include fatty acid synthase, acetyl-CoA carboxylase, glycerol-3-phosphate acyltransferase, and acyl-CoA binding protein. Additionally, SREBPs modulate transcription of stearoyl CoA desaturase-1 and lipoprotein lipase. SREBPs also directly activate transcription of the gene encoding the low density lipoprotein (LDL) receptor, which provides cholesterol and fatty acids through receptor-mediated endocytosis. SREBPs are also implicated in the process of fat cell differentiation and adipose cell gene expression, particularly as transcription factors that can promote adipogenesis in a dominant fashion (reviewed by Spiegelman *et al.*, Cell (1996) 87:377-389).

In high sterol conditions, SREBPs are retained as membrane-bound protein precursors that are kept inactive by virtue of being attached to the nuclear envelope and endoplasmic reticulum (ER) and therefore, excluded from the nucleus. As depicted in Figure 1A, an SREBP in its membrane-bound form has large N-terminal and C-terminal segments facing the cytoplasm and a short loop projecting into the lumen of the organelle. The N-terminal domain is a transcription factor of the basic-helix-loop-helix-leucine zipper (bHLH-Zip) family, and contains an "acid blob" typical of many transcriptional activators. (Brown and Goldstein, Cell (1997) 89:331-340)

The N-terminal acid blob is followed by a basic helix-loop-helix/leucine zipper domain (bHLH-Zip) similar to those found in many other DNA-binding transcriptional regulators. bHLH-Zip domains have two functions: the helix-loop-helix subdomain mediates dimerization, and the basic region binds to specific DNA sequences that include a direct repeat of 5'-PyCAPy-3'. SREBP binds to the sequence 5'-ATCACCCCAC-3' which is known as "sterol regulatory element 1" (SRE-1) and is upstream of the LDL receptor gene.

SREBPs are unique among bHLH-Zip proteins by virtue of the C-terminal domains attached to the bHLH-Zip domain. These include (from – to C-terminus): (1) a hydrophobic membrane-spanning sequence of about 20 amino acids, (2) a hydrophilic stretch of about 31 amino acids that projects into the lumen of the ER, (3) a second hydrophobic membrane-spanning domain of about 20 amino acids, and (4) a C-terminal domain which, in vertebrates, has been determined to be required for sterol regulation of SREBP cleavage.

In low sterol conditions, the acid blob/bHLH-Zip domain of SREBP is released from the membrane after which it is rapidly translocated into the nucleus and binds specific DNA sequences to activate transcription. Two sequential proteolytic cleavages are involved. Referring to Figure 1B, a first protease, referred to as the site 1 protease (S1P) cleaves SREBP at approximately the middle of the lumenal loop. S1P has been cloned from Chinese hamster ovary (CHO) cells (GI (GenBank Identifier No. (hereinafter "GI") 3892203) and a human cell line (GI4506774) (Sakai *et al.*, J. Biol. Chem (1998) 273:5785-5793), and encodes a membrane bound glycoprotein of 1052 amino acids with subtilisin-like sequence features.

After cleavage at site 1, a second protease (the site 2 protease, S2P) cleaves the N-terminal fragment and releases the mature N-terminal domain into the cytosol, from which it rapidly enters the nucleus, apparently with a portion of the transmembrane domain still attached at the C-terminus. Mature, transcriptionally active SREBP is rapidly degraded in a proteosome-dependent process. This combination of proteolytic processing and rapid

turnover allows the SREBP system to rapidly respond to changes in cellular membrane components. S2P homologues have been identified in both vertebrates and invertebrates and have been cloned from human cells and hamster cells (Rawson *et al.*, Molec Cell (1997) 1:47-57). It is a membrane protein containing an HEXXH sequence characteristic of zinc metalloproteases. This family of proteins has high hydrophobicity throughout the amino acid sequence, suggesting the existence of several membrane-spanning regions.

A third component of the processing system for SREBPs is called SREBP Cleavage Activating Protein (SCAP). SCAP is a large transmembrane protein that activates S1P in low-sterol conditions. The N-terminal 730 amino acids have alternating hydrophobic and hydrophilic sequences which are predicted to form up to eight membrane spanning sequences separated by short hydrophilic stretches. This domain is strikingly similar to a domain of HMG CoA reductase (Hua *et al.*, Cell (1996) 87:415-426) which is necessary to impart sterol regulation. In low sterol conditions, HMG-CoA reductase is quite stable, but when sterols are added the enzyme is rapidly degraded. It is believed that the membrane-spanning domain in SCAP, like its counterpart in HMG CoA reductase, can sense the levels of sterol in the ER membrane, either directly or indirectly.

The C-terminal domain of SCAP is hydrophilic and is made up of about 550 amino acids organized into four WD repeats. Recent work has demonstrated that these WD repeats bind directly to the C-terminal regulatory domain of SREBP suggesting that SCAP and SREBP are part of a stable complex in the membrane of the ER (Sakai *et al.*, *supra*). It is likely that S1P and perhaps S2P are also part of the complex since SCAP is essential for activation of S1P activity. This SREBP processing complex is depicted in Figure 2.

The involvement of the SREBP pathway in the regulation of cholesterol metabolism is of interest not only because excess blood cholesterol can lead to atherosclerosis, but also because there seem to be parallels between the processing of SREBPs and the processing of β-amyloid precursor protein which has been implicated in Alzheimer's disease (Brown and Goldstein, *supra*). To date, the SREBP pathway has been studied primarily using mammalian cell culture, by the isolation of mutant cells that are defective in regulation of cholesterol metabolism or intracellular cholesterol trafficking. The mutants can then serve as hosts for cloning genes by functional complementation. This has led to the molecular cloning of the S1P, S2P and SCAP genes (Rawson *et al.*, *supra*; Hua *et al.*, *supra*; and Goldstein *et al.*, US Pat. Nos. 5,527,690 and 5,891,631).

Some SREBP pathway genes have been identified in invertebrates. The isolation of a *Drosophila* SREBP, referred to as "HLH106" (GI079656) has been described (Theopold *et al.*, Proc. Natl. Acad. Sci., USA, (1996) 93(3):1195-1199). An expressed sequence tag (EST) from *C. elegans* which has homology to S2P is described by Rawson *et al.*, *supra*

and is listed in GenBank (GI1559384). Additionally, GenBank has listed a protein predicted from the *C. elegans* genome as having HMG-CoA reductase homology (GI3875380).

SUMMARY OF THE INVENTION

The use of invertebrate model organism genetics can greatly facilitate the elucidation of biochemical pathways, and the identification of molecules that can modulate such pathways. Accordingly, it is an object of the invention to provide invertebrate nucleic acids and polypeptides involved in the SREBP pathway. It is also an object of the invention to provide invertebrate model organisms, including novel mutant phenotypes, for the study of lipid metabolism in general, and more particularly, for the elucidation of the SREBP pathway. It is a further object of the invention to provide methods for screening molecules that modulate lipid metabolism and/or the function of genes and proteins involved in the SREBP pathway.

These and other objects are provided by flies and nematodes that have been genetically modified to express or misexpress an SREBP pathway gene, for example using transposon mutagenesis, RNA interference, chemical mutagenesis, or other genetic techniques. In certain embodiments, expression of the SREBP pathway protein is driven by a heterologous promoter that is tissue-specific, developmentally-specific, or inducible, so that the effects of the expression or mis-expression can be observed in specific tissues, at certain developmental stages, or at specified times, respectively. Additionally, the SREBP pathway protein may be linked to one or more selectable markers that allows detection of expression. Typically, the expression of the SREBP pathway protein results in an identifiable phenotype. In the case of nematodes, the invention provides novel methods for the *in vivo* measurement of lipid content using BODIPY-fatty acid conjugates. The animal models can be used in genetic screens to identify other genes involved in lipid metabolism. They can also be used for screening small molecule libraries directly on whole organisms for possible therapeutic or pesticide use.

The invention also provides novel isolated nucleic acids (SEQ ID NOs:1, 3, and 5) and the SREBP pathway proteins encoded thereby (SEQ ID NOs:2, 4, and 6, respectively), as well as derivatives and fragments thereof. Methods are provided for constructing vectors containing the isolated nucleic acids. Such vectors can be used for making the animal models of the invention. They can also be introduced into host cells to be used for a variety of purposes including two-hybrid screening assays, production of SREBP pathway proteins, screening small molecules that affect lipid synthesis or metabolism, etc.

BRIEF DESCRIPTION OF THE DRAWINGS

Figs. 1A and 1B depict the inactive, membrane-bound form of SREBP (Fig. 1A) and the two-step proteolytic cleavage which activates SREBP in low sterol conditions (Fig. 1B).

Fig. 2 depicts the presumed interactions between SREBP, SCAP, S1P and S2P in the SREBP processing complex.

Figs. 3A-3E show a cDNA sequence that encodes *C. elegans* SREBP (SEQ ID NO:1).

Fig. 4 shows the predicted amino acid sequence (SEQ ID NO:2) of the polypeptide encoded by the *C. elegans* SREBP gene.

Figs. 5A-5C show a cDNA sequence that encodes *Drosophila* S2P (SEQ ID NO:3).

Fig. 6 shows the predicted amino acid sequence (SEQ ID NO:4) of the polypeptide encoded by the *Drosophila* S2P gene.

Figs. 7A-7F show the cDNA sequence that encodes *Drosophila* SCAP (SEQ ID NO:5).

Fig. 8 shows the predicted amino acid sequence (SEQ ID NO:6) of the polypeptide encoded by the *Drosophila* SCAP gene.

Figs. 9A-9E show the nucleic acid sequence encoding *Drosophila* SREBP (GI079656; SEQ ID NO:7).

Fig. 10 shows the predicted amino acid sequence (SEQ ID NO:8) of *Drosophila* SREBP.

DETAILED DESCRIPTION OF THE INVENTION

The use of invertebrate model organism genetics and related technologies can greatly facilitate the elucidation of biological pathways (Scangos, Nat. Biotechnol. (1997) 15:1220-1221; Margolis and Duyk, Nat. Biotechnol. (1998) 16:311). Of particular use are the insect and nematode model organisms, *Drosophila melanogaster*, and *C. elegans*. An extensive search for SREBP pathway nucleic acids and their encoded proteins in *C. elegans* and *Drosophila melanogaster* was conducted in an attempt to identify new and useful tools for probing the function and regulation of the SREBP pathway. Novel SREBP pathway nucleic acids and their encoded proteins are identified herein. As used in this description, the term "SREBP pathway nucleic acid" refers to a nucleic acid that encodes any one of SREBP, SCAP, S1P, and S2P. The newly identified SREBP pathway nucleic acids have led to the discovery of several mutant phenotypes that can be used to study the pathways involved in lipid and fatty acid metabolism. The use of invertebrate model organisms such as *Drosophila melanogaster* and *C. elegans* for analyzing the expression and

mis-expression of SREBP pathway proteins has great advantages over the traditional approach of using mammalian cell culture due to the ability to rapidly carry out large-scale, systematic genetic screens as well as the ability to screen small molecule libraries directly on whole organisms. Thus, the invention provides a superior approach for identifying other components involved in the synthesis, activation, control and turnover of SREBP pathway proteins. Systematic genetic analysis of the SREBP pathway using invertebrate model organisms can lead to the identification of new drug targets, therapeutic agents, diagnostics and prognostics useful in the treatment of disorders associated with lipid metabolism. Additionally, use of these invertebrate model organisms could lead to the identification and validation of pesticide targets directed to components of the SREBP pathway.

The details of the conditions used for the identification and/or isolation of each novel SREBP pathway nucleic acid and protein are described in the Examples section below. Various non-limiting embodiments of the invention and applications and uses of these novel *C. elegans* and *Drosophila melanogaster* SREBP pathway genes and proteins are discussed in the following sections. The entire contents of all references cited herein are incorporated by reference in their entireties for all purposes. Additionally, the citation of a reference in the preceding background section is not an admission of prior art against the claims appended hereto.

Nucleic acids of the SREBP pathway

The invention relates generally to nucleic acid sequences of the SREBP pathway, and more particularly SREBP pathway nucleic acid sequences of *C. elegans* and *Drosophila melanogaster*, and methods of using these sequences. As described in the Examples below, the present invention provides a nucleic acid sequence (SEQ ID NO:1) that was isolated from *C. elegans* and encodes an SREBP homologue referred to herein as "ceSREBP". The invention also provides nucleic acid sequences that were isolated from *Drosophila melanogaster* and encode homologues of S2P (dS2P; SEQ ID NO:3) and SCAP (dSCAP; SEQ ID NO:5). In addition to the fragments and derivatives of SEQ ID NOs 1, 3, and 5, as described in detail below, the invention includes the reverse complements thereof. Also, the subject nucleic acid sequences, derivatives and fragments thereof may be RNA molecules comprising the nucleotide sequence of any one of SEQ ID NOs 1, 3, and 5 (or derivative or fragment thereof) wherein the base U (uracil) is substituted for the base T (thymine). The DNA and RNA sequences of the invention can be single- or double-stranded. Thus, the term "nucleic acid sequence", as used herein, includes the reverse complement, RNA equivalent, DNA or RNA double-stranded sequences, and DNA/RNA hybrids of the sequence being described, unless otherwise indicated explicitly or by context.

Fragments of these sequences can be used for a variety of purposes, for example, as nucleic acid hybridization probes and replication/amplification primers. Certain "antisense" fragments, i.e. that are reverse complements of the sequences set forth in any one of SEQ ID NOs: 1, 3, and 5, have utility in inhibiting the function of SREBP pathway proteins. The fragments are of length sufficient to specifically hybridize with the corresponding SEQ ID NO 1, 3, or 5. In particular, the invention provides fragments of at least 12, preferably at least 24, more preferably at least 36, and more preferably at least 96 contiguous nucleotides of any one of SEQ ID NOs: 1, 3, and 5. In some embodiments, fragments of at least 200 or 500 nucleotides may be preferred. When the fragments are flanked by other nucleic acid sequences, the total length of the combined nucleic acid sequence is less than 15 kb, and preferably less than 10kb, more preferably less than 2 kb, and in some embodiments, more preferably less than 500 bases.

Preferred fragments of ceSREBP (SEQ ID NO:1) include those having at least 535 contiguous nucleotides of SEQ ID NO:1, and more preferably at least 540 nucleotides. In another embodiment of the invention, a fragment contains approximately residues 1090 to 1290 of SEQ ID NO:1, which encodes a bHLH-Zip domain. Other preferred fragments comprise any one of the following contiguous sequences of SEQ ID NO:1: nucleotides 1-85, 70-90, 76-218, 203-223, 208-528, 513-533, 517-637, 623-643, 626-1058, 1043-1063, 1048-1293, 1279-1299, 1277-1486, 1473-1493, 1477-2016, 2002-2022, 2004-2413, 2399-2419, 2404-2641, 2627-2647, 2632-2795, 2781-3001, 2786-3156, 3142-3162, and 3147-3397.

Preferred fragments of dS2P (SEQ ID NO:3) include those having at least 1226 contiguous nucleotides of SEQ ID NO:3, and more preferably at least 1231 nucleotides. Other preferred fragments comprise any one of the following contiguous sequences of SEQ ID NO:3: nucleotides 5-296, 281-301, 287-734, 719-739, and 725-1958.

Preferred fragments of dSCAP (SEQ ID NO:5) include those having at least 2274 contiguous nucleotides of SEQ ID NO:5, and more preferably at least 2279 nucleotides. Other preferred fragments comprise any one of the following contiguous sequences of SEQ ID NO:5: nucleotides 1-160, 150-170, 151-544, 529-549, 526-719, 704-724, 711-2988, 2974-3004, 2981-3191, 3177-3197, 3182-3546, 3532-3552 and 3537-3765.

Additionally, fragments of any of the foregoing sequences that are double-stranded RNA (dsRNA) molecules have utility in RNA interference (RNAi) studies, as described in more detail below, where model organisms exhibiting loss-of-function phenotype are generated. Typically, dsRNA molecules for RNAi studies are from about 200 to 2000 bp, and are preferably 600-900 bp in size.

The subject nucleic acid sequences may consist solely of any one of SEQ ID NOs:1, 3, and 5, or fragments thereof. Alternatively, the subject nucleic acid sequences and fragments thereof may be joined to other components such as labels, peptides, agents that facilitate transport across cell membranes, hybridization-triggered cleavage agents or intercalating agents. The subject nucleic acid sequences and fragments thereof may also be joined to other nucleic acid sequences (i.e. they may comprise part of larger sequences) and are of synthetic/non-natural sequences and/or are isolated, i.e. unaccompanied by at least some of the material with which it is associated in its natural state. Preferably, the isolated nucleic acids constitute at least about 0.5%, and more preferably at least about 5% by weight of the total nucleic acid present in a given fraction, and are preferably recombinant, meaning that they comprise a non-natural sequence or a natural sequence joined to nucleotide(s) other than that which it is joined to on a natural chromosome.

The invention also provides derivative nucleic acid sequences which hybridize to the nucleic acid sequence of any one of SEQ ID NOs:1, 3, and 5 under stringency conditions such that each hybridizing derivative nucleic acid is related to the subject nucleic acid by a certain degree of sequence identity. In a specific embodiment, the derivative nucleic acid hybridizes to the reverse complement of SEQ ID NO:1, 3 or 5 and has the antigenicity of a polypeptide encoded by SEQ ID NO:2, 4, or 6, respectively. The temperature and salt concentrations at which hybridizations are performed have a direct effect on the results that are obtained. With "stringent" or "high stringency" conditions, a denaturing agent, such as formamide, is used during hybridization. The formamide is typically used at 25% to 50% (v/v) in a buffered diluent comprising 1X to 6X SSC (1X SSC is 150 mM NaCl and 15mM sodium citrate; SSPE may be substituted for SSC, 1X SSPE is 150mM NaCl, 10 mM Na H_2PO_4, and 1.25 mM EDTA, pH7.4). The hybridization temperature is typically about 42°C. High stringency conditions also employ a wash buffer with low ionic strength, such as 0.1X to about 0.5X SSC, at relatively high temperature, typically greater than about 55°C up to about 70°C. Moderately stringent conditions typically use 0% to 25% formamide in 1X to 6X SSC, and use reduced hybridization temperatures, usually in the range of about 27°C to about 40°C. The wash buffer can have increased ionic strength, *e.g.* about 0.6X to about 2X SSC, and is used at reduced temperatures, usually from about 45°C to about 55°C. With "non-stringent" or "low stringency" hybridization conditions, the hybridization buffer is the same as that used for moderately stringent or high stringency, but does not contain a denaturing agent. A reduced hybridization temperature is used, typically in the range of about 25°C to about 30°C. The wash buffer has increased ionic strength, usually around 2X to about 6X SSC, and the wash temperature is in the range of about 35°C to about 47°C. Procedures for nucleic acid hybridizations are well-known in the art

(Ausubel *et al.*, Current Protocols in Molecular Biology (1995) Wiley Interscience Publishers; Sambrook *et al.*, Molecular Cloning: A Laboratory Manual (1989) Cold Spring Harbor Press, New York; Shilo and Weinberg, Proc. Natl. Acad. Sci. U.S.A. (1981) 78:6789-6792).

In a specific embodiment of the invention, nucleic acids are provided that are capable of hybridizing to any one of SEQ ID NOs:1, 3, and 5, or the above-specified fragments thereof, under any one of the hybridization conditions listed in Table 1. Hybridization conditions 8-10, as listed in Table 1, are generally considered "high stringency" conditions; conditions 4-7 are generally considered "moderately stringent", and conditions 1-3 are considered "non-stringent".

TABLE I

Condition #	Hybridization Buffer	Hybridization Temp.	Wash Buffer	Wash Temp.
1	6X SSC / 0% formamide	25°C	4X SSC	35°C
2	6X SSC / 0% formamide	25°C	4X SSC	40°C
3	6X SSC / 0% formamide	27°C	4X SSC	47°C
4	6X SSC / 0% formamide	34°C	2X SSC	45°C
5	6X SSC / 0% formamide	40°C	0.8X SSC	45°C
6	3X SSC / 0% formamide	40°C	0.6X SSC	50°C
7	1X SSC / 0% formamide	40°C	0.6X SSC	55°C
8	6X SSC / 25% formamide	42°C	0.5X SSC	60°C
9	2X SSC / 25% formamide	42°C	0.4X SSC	65°C
10	1X SSC / 25% formamide	42°C	0.3X SSC	70°C

Condition #1 shown in Table 1 is designed to isolate nucleic acids having at least about 50% sequence identity with the target nucleic acid (with % identity calculated as described below). With each subsequent condition, the stringency is such that the isolated nucleic acid has a sequence identity of at least 5% greater than what would be isolated by using the next lower condition number. Thus, for example, condition #2 is designed to isolate nucleic acids having at least about 55% sequence identity with the target nucleic acid, and conditions #9 and #10 are designed to isolate nucleic acids having at least about 90% and 95% sequence identity, respectively, to the target nucleic acid. Preferably, each hybridizing derivative nucleic acid has a length that is at least 30% of the length of the subject nucleic acid sequence described herein to which it hybridizes. More preferably, the hybridizing

nucleic acid has a length that is at least 50%, still more preferably at least 70%, and most preferably at least 90% of the length of the subject nucleic acid sequence to which it hybridizes.

As used herein, "percent (%) nucleic acid sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides in the candidate derivative nucleic acid sequence identical with the nucleotides in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410; http://blast.wustl.edu/blast/README.html) with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A % nucleic acid sequence identity value is determined by the number of matching identical nucleotides divided by the sequence length for which the percent identity is being reported. Preferably, derivative nucleic acid sequences of the present invention have at least 70% preferably at least 80%, more preferably at least 85%, still more preferably at least 90%, and most preferably at least 95% sequence identity with any one of SEQ ID NOs:1, 3, and 5. In some preferred embodiments, the derivative nucleic acid encodes a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs:2, 4, and 6, or a functionally active fragment thereof.

A derivative of the subject nucleic acid sequence, or fragment thereof, may comprise 100% sequence identity with the subject nucleic acid sequence, but be a derivative thereof in the sense that it has one or more modifications at the base or sugar moiety, or phosphate backbone. Examples of modifications are well known in the art (Bailey J.E. Ullmann's Encyclopedia of Industrial Chemistry (1998), 6th ed. Wiley and Sons). Such derivatives may be used to provide modified stability or any other desired property.

Another type of derivative of the subject nucleic acid sequences includes corresponding humanized sequences. A humanized nucleic acid sequence is one in which one or more codons has been substituted with a codon that is more commonly used in human genes. The following list shows, for each amino acid, the calculated codon frequency (number in parentheses) in humans genes for 1000 codons (Wada *et al.*, Nucleic Acids Research (1990) 18(Suppl.):2367-2411):

Human codon frequency per 1000 codons:

```
ARG: CGA (5.4),  CGC (11.3), CGG (10.4), CGU (4.7),  AGA (9.9),  AGG (11.1)
LEU: CUA (6.2),  CUC (19.9), CUG (42.5), CUU (10.7), UUA (5.3),  UUG (11.0)
SER: UCA (9.3),  UCC (17.7), UCG (4.2),  UCU (13.2), AGC (18.7), AGU (9.4)
```

```
THR: ACA (14.4), ACC (23.0), ACG (6.7),  ACU (12.7)
PRO: CCA (14.6), CCC (20.0), CCG (6.6),  CCU (15.5)
ALA: GCA (14.0), GCC (29.1), GCG (7.2),  GCU (19.6)
GLY: GGA (17.1), GGC (25.4), GGG (17.3), GGU (11.2)
VAL: GUA (5.9),  GUC (16.3), GUG (30.9), GUU (10.4)
LYS: AAA (22.2), AAG (34.9)
ASN: AAC (22.6), AAU (16.6)
GLN: CAA (11.1), CAG (33.6)
HIS: CAC (14.2), CAU (9.3)
GLU: GAA (26.8), GAG (41.4)
ASP: GAC (29.0), GAU (21.7)
TYR: UAC (18.8), UAU (12.5)
CYS: UGC (14.5), UGU (9.9)
PHE: UUU (22.6), UUC (15.8)
ILE: AUA (5.8),  AUC (24.3), AUU (14.9)
MET: AUG (22.3)
TRP: UGG (13.8)
TER: UAA (0.7),  AUG (0.5),  UGA (1.2)
```

Thus, an SREBP pathway nucleic acid sequence in which the glutamic acid codon, GAA has been replaced with the codon GAG, which is more commonly used in human genes, is an example of a humanized SREBP pathway nucleic acid sequence. A detailed discussion of the humanization of nucleic acid sequences is provided in U.S. Pat. No. 5,874,304 to Zolotukhin *et al.*

Isolation, production, and expression of nucleic acids of the SREBP pathway

Nucleic acid encoding the amino acid sequence of any one of SEQ ID NOs:2, 4, and 6, may be obtained from an appropriate cDNA library prepared from any eukaryotic species that encodes SREBP pathway proteins such as vertebrates, preferably mammalian (*e.g.* primate, porcine, bovine, feline, equine, and canine species, etc.) and invertebrates, such as arthropods, particularly insects species (preferably *Drosophila melanogaster*) and arachnids, and nematodes (preferably *C. elegans*). An expression library can be constructed using known methods. For example, mRNA can be isolated to make cDNA which is ligated into a suitable expression vector for expression in a host cell into which it is introduced. Various screening assays can then be used to select for the gene or gene product (*e.g.* oligonucleotides of at least about 20 to 80 bases designed to identify the gene of interest, or labeled antibodies that specifically bind to the gene product).

Polymerase chain reaction (PCR) can also be used to isolate nucleic acids of the SREBP pathway where oligonucleotide primers representing fragmentary sequences of interest amplify RNA or DNA sequences from a source such as a genomic or cDNA library (as described by Sambrook *et al., supra*). Additionally, degenerate primers for amplifying homologues from any species of interest may be used. Once a PCR product of appropriate size and sequence is obtained, it may be cloned and sequenced by standard techniques, and utilized as a probe to isolate a complete cDNA or genomic clone.

Fragmentary sequences of SEQ ID NOs 1, 3 and 6 may be synthesized by known methods. For example, oligonucleotides may be synthesized using an automated DNA synthesizer available from commercial suppliers (*e.g.* Biosearch, Novato, CA; Perkin-Elmer Applied Biosystems, Foster City, CA). Antisense RNA sequences can be produced intracellularly by transcription from an exogenous sequence, *e.g.* from vectors that contain antisense SREBP pathway nucleic acid sequences. Newly generated sequences may be identified and isolated using standard methods.

An isolated SREBP pathway nucleic acid sequence can be inserted into any appropriate cloning vector, for example bacteriophages such as lambda derivatives, or plasmids such as PBR322, pUC plasmid derivatives and the Bluescript vector (Stratagene, San Diego, CA). Recombinant molecules can be introduced into host cells via transformation, transfection, infection, electroporation, *etc.* The transformed cells can be cultured to generate large quantities of the SREBP pathway nucleic acid. Suitable methods for isolating and producing the subject nucleic acid sequences are well-known in the art (Sambrook *et al.*, *supra*; Glover (ed.), DNA Cloning: A Practical Approach, Vol. 1, 2, 3, 4, (1995) MRL Press, Ltd., Oxford, U.K.).

The nucleotide sequence coding an SREBP pathway protein or a functionally active fragment or derivative thereof, can be inserted into any appropriate expression vector for the transcription and translation of the inserted protein-coding sequence. Alternatively, the necessary transcriptional and translational signals can be supplied by the native SREBP pathway gene and/or its flanking regions. A variety of host-vector systems may be utilized to express the protein-coding sequence such as mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.* baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. Expression of an SREBP pathway protein may be controlled by a suitable promoter/enhancer element. In addition, a host cell strain may be selected which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired.

In a specific embodiment, a vector is used that comprises a promoter operably linked to an SREBP pathway gene nucleic acid, one or more origins of replication, and optionally, one or more selectable markers (*e.g.* thymidine kinase activity, resistance to antibiotics, *etc.*) so that expression of the gene product can be detected. Alternatively, recombinant expression vectors can be identified by assaying for the expression of the SREBP pathway gene product based on the physical or functional properties of the SREBP pathway protein in *in vitro* assay systems (*e.g.* immunoassays).

In specific embodiments, the SREBP pathway protein, fragment, or derivative may be expressed as a fusion, or chimeric protein product (i.e. it is joined via a peptide bond to a heterologous protein sequence of a different protein). Such a chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame (using methods known in the art) and expressing the chimeric product. Alternatively, such a chimeric product may be made by protein synthetic techniques, *e.g.* by use of a peptide synthesizer.

Once a recombinant which expresses the SREBP pathway gene sequence is identified, the gene product can be isolated using standard methods (*e.g.* ion exchange, affinity, and sizing column chromatography; centrifugation; differential solubility). The amino acid sequence of the protein can be deduced from the nucleotide sequence of the chimeric gene contained in the recombinant. As a result, the protein can be synthesized by standard chemical methods known in the art (Hunkapiller *et al.*, Nature (1984) 310:105-111). Alternatively, native SREBP-pathway proteins can be purified from natural sources, by standard methods (*e.g.* immunoaffinity purification).

SREBP pathway proteins

The invention provides SREBP pathway proteins that comprise or consist of an amino acid sequence of any one of SEQ ID NOs: 2, 4, and 6, or fragments or derivatives thereof. Compositions comprising these proteins may consist essentially of the SREBP pathway proteins. Alternatively, the SREBP pathway proteins may be a component of a composition that comprises other components (*e.g.* a diluent such as saline, a pharmaceutically acceptable carrier or excipient, a culture medium, carriers used in pesticide formulations, etc.).

Typically, a derivative of an SREBP pathway protein will share a certain degree of sequence identity or sequence similarity with any one of SEQ ID NOs 2, 4, and 6, or a fragment thereof. As used herein, "percent (%) amino acid sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of amino acids in the candidate derivative amino acid sequence identical with the amino acid in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, *supra*) using the same parameters discussed above for derivative nucleic acid sequences. A % amino acid sequence identity value is determined by the number of matching identical amino acids divided by the sequence length for which the percent identity is being reported. Preferably, derivative amino acid sequences of the present invention have at least 80%, preferably at

least 85%, more preferably at least 90%, and most preferably at least 95% sequence identity with any contiguous stretch of at least 20 amino acids, preferably at least 25 amino acids, and more preferably at least 30 amino acids of any one of SEQ ID NOs 2, 4, and 6. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity described above, but including conservative amino acid substitutions in additional to identical amino acids in the computation. A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids arginine, lysine and histidine; interchangeable acidic amino acids aspartic acid and glutamic acid; and interchangeable small amino acids alanine, serine, threonine, methionine, and glycine.

A preferred derivative of ceSREBP consists of or comprises an amino acid sequence that has at least 55%, preferably at least 66%, and more preferably, at least 65% sequence identity with amino acid residues 335-428 of SEQ ID NO:2 (i.e. the bHLH-Zip domain). Other preferred derivatives of ceSREBP consist of or comprise an amino acid sequence that shares at least 75% similarity, preferably at least 80% similarity, and more preferably, at least 85% similarity with amino acid residues 335-428 of SEQ ID NO:1. Preferably, such derivatives share antigenicity with amino acid residues 335-428 of SEQ ID NO:1.

The invention also provides proteins having amino acid sequences that consist of or comprise a fragment of any one of SEQ ID NOs 2, 4, and 6. The fragments usually have at least 10, preferably at least 12, and more preferably at least 15 contiguous amino acids of any one of SEQ ID NOs 2, 4, and 6. A preferred fragment of ceSREBP contains at least 8, preferably at least 10, and more preferably at least 12 contiguous amino acids of residues 335 to 428 of SEQ ID NO:2.

Preferably the fragment or derivative of the SREBP pathway protein is "functionally active" meaning that the SREBP pathway protein derivative or fragment exhibits one or more functional activities associated with a full-length, wild-type SREBP pathway protein comprising the amino acid sequence of any one of SEQ ID NOs:2, 4, and 6. As an example, functionally active SREBP pathway protein fragments or derivatives include polypeptides that have the antigenicity of the SREBP pathway protein such that they can be used in immunoassays, for immunization, for inhibition of SREBP pathway activity, *etc.* As another example, a fragment or derivative of SREBP may be considered functionally active if it binds a regulatory DNA element of an appropriate target gene such as the SRE-1

sequence. S2P may be considered functionally active if it cleaves SREBP at site 2 (as depicted in Fig. 1B), etc. A fragment or derivative of SCAP may be considered functionally active it is capable of binding to the C-terminal regulatory domain of SREBP. Fragments or derivatives of SREBP pathway proteins can be tested for functional activity by various procedures known in the art. In a preferred method which is described in detail below, a model organism, such as an insect (*e.g. D. melanogaster*) or worm (*e.g. C. elegans*), or other model system, is used in genetic studies to assess the phenotypic effect of a fragment or derivative (i.e. mutant). As used herein, functionally active fragments also include polypeptides that are lacking one or more structural or functional domains of an SREBP pathway protein. Examples of such domains include transmembrane domains, cytosolic domains, lumenal domains, regulatory domains, etc. Thus, for example, an SREBP polypeptide lacking the N-terminal acidic region and/or the C-terminal regulatory region, is considered a functionally-active fragment.

The SREBP pathway derivatives of the invention can be produced by various methods known in the art. The manipulations which result in their production can occur at the gene or protein level. For example, a cloned SREBP pathway gene sequence can be cleaved at appropriate sites with restriction endonuclease(s) (Wells *et al.*, Philos. Trans. R. Soc. London SerA (1986) 317:415), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*, and expressed to produce the desired derivative. Alternatively, an SREBP pathway gene can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or to form new restriction endonuclease sites or destroy preexisting ones, to facilitate further *in vitro* modification. A variety of mutagenesis techniques are known in the art such as chemical mutagenesis, *in vitro* site-directed mutagenesis (Carter *et al.*, Nucl. Acids Res. (1986) 13:4331), use of TAB® linkers (available from Pharmacia and Upjohn, Kalamazoo, MI), *etc.*

At the protein level, manipulations include post translational modification, *e.g.* glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, *etc.* Any of numerous chemical modifications may be carried out by known technique (*e.g.* specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, $NaBH_4$, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, *etc.*). Derivative proteins can also be chemically synthesized by use of a peptide synthesizer, for example to introduce nonclassical amino acids or chemical amino acid analogs as substitutions or additions into the SREBP pathway protein sequence.

Chimeric or fusion proteins can be made comprising an SREBP pathway protein or fragment thereof (preferably consisting of at least a domain or motif of the SREBP pathway protein, or at least 6, and preferably at least 10 amino acids of the SREBP pathway protein) joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein. Such a chimeric protein can be produced by any known method, including: recombinant expression of a nucleic acid encoding the protein (comprising an SREBP pathway-coding sequence joined in-frame to a coding sequence for a different protein); ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other in the proper coding frame, and expressing the chimeric product; and protein synthetic techniques, *e.g.* by use of a peptide synthesizer.

Antibodies to SREBP pathway proteins

SREBP pathway proteins, including functional derivatives and fragments thereof (*e.g.* an SREBP pathway protein encoded by a sequence of any one of SEQ ID NOs:2, 4, and 6, or a subsequence thereof) may be used as an immunogen to generate monoclonal or polyclonal antibodies and antibody fragments or derivatives (*e.g.* chimeric, single chain, Fab fragments). For example, antibodies to a particular domain of an SREBP pathway protein may be desired (*e.g.* an SRE binding domain). In a specific embodiment, fragments of an SREBP pathway protein identified as hydrophilic are used as immunogens for antibody production using art-known methods. Various known methods for antibody production can be used including cell culture of hybridomas; production of monoclonal antibodies in germ-free animals (PCT/US90/02545); the use of human hybridomas (Cole *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1983) 80:2026-2030; Cole *et al.*, in Monoclonal Antibodies and Cancer Therapy (1985) Alan R. Liss, pp. 77-96), and production of humanized antibodies (Jones *et al.*, Nature (1986) 321:522-525; US Pat. No. 5,530,101).

Molecules which interact with SREBP pathway proteins

The present invention provides methods of identifying or screening for molecules, such as proteins or other compounds, which interact with SREBP pathway proteins, or derivatives, or fragments thereof. Assays to find interacting proteins can be performed by any method known in the art, for example, immunoprecipitation with an antibody that binds to the protein in a complex followed by analysis by size fractionation of the immunoprecipitated proteins (*e.g.* by denaturing or nondenaturing polyacrylamide gel electrophoresis), Western analysis, non-denaturing gel electrophoresis, *etc.* A preferred method for identifying interacting proteins is a two hybrid assay system or variation thereof

(Fields and Song, Nature (1989) 340:245-246; U.S. Patent No. 5,283,173; for review see Brent and Finley, Annu. Rev. Genet. (1997) 31:663-704).

The most commonly used two-hybrid screen system is performed using yeast. All systems share three elements: 1) a gene that directs the synthesis of a "bait" protein fused to a DNA binding domain; 2) one or more "reporter" genes having an upstream binding site for the bait, and 3) a gene that directs the synthesis of a "prey" protein fused to an activation domain that activates transcription of the reporter gene. For the screening of proteins that interact with SREBP pathway proteins, the "bait" is preferably an SREBP pathway protein having an amino acid sequence of any one of SEQ ID NOs:2, 4, and 6 (or derivative or fragment thereof), expressed as a fusion protein to a DNA binding domain. Because most two-hybrid systems are engineered to enter the nucleus and activate transcription, transmembrane portions of proteins can interfere with proper association, folding, and nuclear transport of bait or prey segments (Ausubel *et al.*, *supra*; Allen *et al.*, Trends Biochem. Sci. (1995) 20:511-516). Therefore, the "bait" is preferably an SREBP pathway protein derivative or a fragment that lacks transmembrane domains. The "prey" protein is a protein to be tested for ability to interact with the bait, and is expressed as a fusion protein to a transcription activation domain. In one embodiment, the prey proteins can be obtained from recombinant biological libraries expressing random peptides.

The bait fusion protein can be constructed using any suitable DNA binding domain. In a preferred system, the bait contains DNA binding and dimerization domains of the E. coli LexA repressor protein. LexA binds tightly to several different operators, and carries a dimerization domain at its C terminus. In another preferred system, the bait contains residues 1-147 of the yeast GAL4 protein which binds tightly to appropriate DNA binding sites, localizes fused proteins to the nucleus, and directs dimerization (Bartel *et al.*, BioTechniques (1993) 14:920-924, Chasman *et al.*, Mol. Cell. Biol. (1989) 9:4746-4749; Ma *et al.*, Cell (1987) 48:847-853; Ptashne *et al.*, Nature (1990) 346:329-331).

The prey fusion protein can be constructed using any suitable activation domain such as GAL4, VP-16, etc. In various embodiments the preys contain useful moieties such as nuclear localization signals (Ylikomi *et al.*, EMBO J. (1992) 11:3681-3694; Dingwall and Laskey, TIBS (1991) 16:479-481) or epitope tags (Allen *et al.*, *supra*) to facilitate isolation of the encoded proteins. Activation tagged proteins also differ in whether they are expressed constitutively, or conditionally. In a preferred embodiment , the prey is conditionally expressed, allowing the transcription phenotypes obtained in selections (or "hunts") for interactors to be ascribed to the synthesis of the tagged protein, thus reducing the number of false positive cells that grow because their reporters are aberrantly transcribed.

Any reporter gene can be used that has a detectable phenotype. In various specific embodiments, some reporter genes allow cells expressing them to be selected by growth on appropriate medium (*e.g.* HIS3, LEU2 described by Chien *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1991) 88:9572-9582; and Gyuris *et al.*, Cell (1993) 75:791-803). Others allow cells expressing them to be visually screened such as LacZ and GFP (Chien *et al.*, *supra*; and http:/www.bio101.com). Reporters also differ in the number and affinity of upstream binding sites (*e.g.* lexA operators) for the bait, and in the position of these sites relative to the transcription start point. Finally, reporter genes differ in the number of molecules of the reporter gene product necessary to score the phenotype. These differences affect the strength of the protein interactions the reporters can detect. Thus, for example, one or more tandem copies (*e.g.* four or five copies) of the appropriate DNA binding site can be introduced upstream of the TATA box in the desired promoter (*e.g.* in the area of about position -100 to about -400). In a preferred aspect, 4 or 5 tandem copies of the 17 bp UAS (GAL4 DNA binding site) are introduced upstream of the TATA box in the desired promoter, which is upstream of the desired coding sequence for a selectable or detectable marker.

Although the preferred host for two-hybrid screening is the yeast, the host cell in which the interaction assay and transcription of the reporter gene occurs can be any cell, such as mammalian (*e.g.* monkey, mouse, rat, human, bovine), chicken, bacterial, or insect cells. Expression constructs encoding and capable of expressing the binding domain fusion proteins, the transcriptional activation domain fusion proteins, and the reporter gene product(s) are provided within the host cell, by mating of cells containing the expression constructs, or by cell fusion, transformation, electroporation, microinjection, *etc.* The host cell used should not express an endogenous transcription factor that binds to the same DNA site as that recognized by the DNA binding domain fusion population. Also, preferably, the host cell is mutant or otherwise lacking in an endogenous, functional form of the reporter gene(s) used in the assay. Various vectors and host strains for expression of the two fusion protein populations in yeast can be used (U.S. Patent No. 5,1468,614; Bartel *et al.*, Cellular Interactions in Development (1993) Hartley, ed., Practical Approach Series xviii, IRL Press at Oxford University Press, New York, NY, pp. 153-179; and Fields and Sternglanz, Trends In Genetics (1994) 10:286-292. As an example of a mammalian system, interaction of activation tagged VP16 derivatives with a GAL4-derived bait drives expression of reporters that direct the synthesis of Hygromycin B phosphotransferase, Chloramphenicol acetyltransferase, or CD4 cell surface antigen (Fearon *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1992) 89:7958-7962). In another embodiment, interaction of VP16-tagged derivatives with GAL4-derived baits drives the synthesis of SV40 T antigen, which in turn promotes

the replication of the prey plasmid, which carries an SV40 origin (Vasavada *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1991) 88:10686-10690).

False positives arising from transcriptional activation by the DNA binding domain fusion proteins in the absence of a transcriptional activator domain fusion protein can be prevented or reduced by negative selection for such activation within a host cell containing the DNA binding fusion population, prior to exposure to the activation domain fusion population. For example, if such cell contains *URA3* as a reporter gene, negative selection is carried out by incubating the cell in the presence of 5-fluoroorotic acid (5-FOA), which kills self-activating DNA-binding domain hybrids.

In a preferred embodiment, the bait SREBP pathway gene and the prey library of chimeric genes are combined by mating the two yeast strains on solid media for a period of approximately 6-8 hours. Alternatively, the mating can be performed in liquid media. The resulting diploids contain both kinds of chimeric genes, *i.e.*, the DNA-binding domain fusion and the activation domain fusion.

Transcription of the reporter gene can be detected by a linked replication assay, for example, as described by Vasavada *et al.*, *supra*, or using immunoassay methods, preferably as described in Alam and Cook (Anal. Biochem. (1990)188:245-254). The activation of other reporter genes like *URA3*, *HIS3, LYS2, or LEU2* enables the cells to grow in the absence of uracil, histidine, lysine, or leucine, respectively, and hence serves as a selectable marker. Other types of reporters are monitored by measuring a detectable signal. For example, *GFP* and *lacZ* have gene products that are fluorescent and chromogenic, respectively.

After interacting proteins have been identified, the DNA sequences encoding the proteins can be isolated. In one method, the activation domain sequences or DNA-binding domain sequences (depending on the prey hybrid used) are amplified, for example, by PCR using pairs of oligonucleotide primers specific for the coding region of the DNA binding domain or activation domain. Other known amplification methods can be used, such as ligase chain reaction, use of Q replicase, or methods described by Kricka *et al.* (Molecular Probing, Blotting, and Sequencing (1995) Academic Press, New York, Chapter 1 and Table IX).

If a shuttle (yeast to *E. coli*) vector is used to express the fusion proteins, the DNA sequences encoding the proteins can be isolated by transforming the yeast DNA into *E. coli* and recovering the plasmids from *E. coli*. Alternatively, the yeast vector can be isolated, and the insert encoding the fusion protein subcloned into a bacterial expression vector, for growth of the plasmid in *E. coli*.

The two hybrid system can be used for screening candidate molecules that modulate interaction between the SREBP pathway protein and the protein with which it interacts. Briefly, the protein-protein interaction assay can be carried out by assaying for reporter gene activity as described above, except that it is done in the presence of one or more candidate molecules. An increase or decrease in reporter gene activity relative to that present when the one or more candidate molecules are absent indicates that the candidate molecule has an effect on the interacting pair. In a preferred method, inhibition of an interaction is selected when the inhibition is necessary for cell survival (*e.g.* an interaction that activates the *URA3* gene, causing yeast to die in medium containing the chemical 5-fluoroorotic acid (Rothstein, Meth. Enzymol. (1983)101:167-180)). The identification of inhibitors of such interactions can also be accomplished using competitive inhibitor assays.

***In vivo* and *in vitro* models of SREBP pathway gene function and dysfunction**

Both genetically modified animal models (i.e. *in vivo* models), such as *C. elegans* and *Drosophila melanogaster*, and *in vitro* models such as genetically engineered cell lines expressing or mis-expressing SREBP pathway genes, are useful for studying lipid metabolism and disorders associated with abnormal lipid metabolism. Such models that display detectable phenotypes, such as those described in more detail below and in the examples, can be used for the identification and characterization of SREBP pathway genes or other genes of interest and/or phenotypes associated with the mutation or mis-expression of an SREBP pathway protein. The term "mis-expression" as used herein encompasses mis-expression due to gene mutations. Thus, a mis-expressed SREBP pathway protein may be one having an amino acid sequence that differs from wild-type (i.e. it is a derivative of the normal protein). A mis-expressed SREBP pathway protein may also be one in which one or more amino acids have been deleted, and thus is a "fragment" of the normal protein. As used herein, "mis-expression" also includes over-expression (*e.g.* by multiple gene copies), underexpression, and non-expression (*e.g.* by gene knockout or blocking expression that would otherwise normally occur). As used in the following discussion concerning *in vivo* and *in vitro* models, the term "gene of interest" refers to an SREBP pathway gene (i.e. SREBP, SCAP, S1P, and S2P), or any gene involved in regulation or modulation of the SREBP pathway. Such genes may include any gene involved in the biosynthesis or metabolism of cholesterol or fatty acids such as HMG coenzyme A synthase, HMG-CoA reductase, farnesyl diphosphate synthase, squalene synthase, fatty acid synthase, acetyl-CoA carboxylase, glycerol-3-phosphate acyltransferase, acyl-CoA binding protein, stearoyl CoA desaturase-1, lipoprotein lipase, and the LDL receptor.

The *in vivo* and *in vitro* models may be genetically engineered or modified so that they 1) have deletions and/or insertions of one or more SREBP pathway genes, 2) harbor interfering RNA sequences derived from SREBP pathway genes, 3) have had one or more endogenous SREBP pathway genes mutated, and/or 4) contain transgenes for mis-expression of wild-type or mutant forms of such genes. Such genetically modified *in vivo* and *in vitro* models are useful for identification of new genes that are involved in the synthesis, activation, control, etc. of SREBP pathway genes and/or gene products. Further, other genes of interest that are involved in cholesterol and/or fatty acid biosynthesis or metabolism may be identified. The newly identified genes could constitute possible pesticide targets (as judged by animal model phenotypes such as non-viability, block of normal development, defective feeding, defective movement, or defective reproduction). Alternatively, or additionally, they may constitute possible therapeutic targets, particularly in the area of metabolic diseases and disorders, for example, cholesterol synthesis, metabolism, and other fatty acid disorders. The model systems can also be used for testing potential pesticidal or pharmaceutical compounds that interact with the SREBP pathway, for example by administering the compound to the model system using any suitable method (*e.g.* direct contact, ingestion, injection, etc.) and observing any changes in phenotype, for example, changes in lipid content, lethality, etc.

A variety of known expression modification methods can be used to genetically modify the animal models and cell cultures so that they express or mis-express SREBP pathway proteins. Some specific examples include radiation mutagenesis such as X-rays, gamma rays, and ultraviolet radiation; chemical mutagenesis using, for example, ethylmethane sulfonate (EMS), methylmethane sulfonate (MMS), N-ethyl-N-nitrosourea (ENU), triethylmelamine, diepoxyalkanes, ICR-170, or formaldehyde; double-stranded RNA interference; use of peptide and RNA aptamers; transposon mutagenesis and transgene-mediated mis-expression. For some applications, it is useful to use genetic modification techniques that result in inheritable expression or mis-expression patterns such that the progeny of the genetically-modified animals can be studied. Various genetic modification techniques are discussed in more detail below and in the Examples.

Chemical mutagenesis

A commonly-used chemical mutagen for creating loss-of-function mutations is ethyl methanesulfonate (EMS). In *C. elegans*, EMS mutagenesis can result in small deletions at a rate of approximately 13%. Accordingly, there is about a 95% probability of identifying a deletion in a gene of interest by screening 4 x 10^6 EMS-mutagenized genomes. Briefly, several million nematodes are mutagenized with EMS using the procedure described by

Sulston and Hodgkin (The nematode *Caenorhabditis elegans* (1988) Wood, Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 587-606). The mutagenized nematodes are then distributed in small pools in 96-well plates, each pool composed of approximately 400 haploid genomes. A portion of each pool is used to generate a corresponding library of genomic DNA derived from the mutagenized nematodes. The DNA library is screened with a PCR assay to identify pools that carry genomes with deletions of interest. Mutant worms carrying the desired deletions are recovered from the corresponding pools of the mutagenized animals. Although EMS is a preferred mutagen to generate deletions, other mutagens can be used that also provide a significant yield of deletions, such as X-rays, gamma-rays, diepoxybutane, formaldehyde and trimethylpsoralen with ultraviolet light.

Chemical mutagenesis methods, and other methods, for generating loss-of-function mutations in *D. melanogaster* are described by Ashburner (*Drosophila*: A Laboratory Manual (1989) Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press).

RNA-mediated interference with gene expression

RNA-mediated interference (RNAi), is an effective method for generating loss-of-function phenotypes. Loss-of-function phenotypes can be generated by injecting antisense RNA that is partially homologous to a gene of interest into embryos using methods described by Schubiger and Edgar (Methods in Cell Biology (1994) 44:697-713). Another antisense RNA methodology involves expression of an antisense RNA partially homologous to the gene of interest by operably joining a portion of the gene in the antisense orientation to a powerful promoter (such as heat shock gene promoters, or promoters controlled by potent exogenous transcription factors, such as GAL4 and tTA) that can drive the expression of large quantities of the antisense RNA. Antisense RNA-generated loss-of-function phenotypes have been reported for several *Drosophila* genes (LaBonne *et al.*, Dev. Biol. (1989) 136(1):1-16; Schuh and Jackle, Genome (1989) 31(1):422-5; and Geisler *et al.*, Cell (1992) 71(4):613-21).

Loss-of-function phenotypes can also be generated by cosuppression methods where a sense strand RNA corresponding to a partial segment of the gene of interest is injected into the animal (Bingham, Cell (1997) 90(3):385-7; Smyth, Curr. Biol. (1997) 7(12):793-5; Que and Jorgensen, Dev. Genet. (1998) 22(1):100-9; and Pal-Bhadra *et al.*, Cell (1997) 90(3):479-90).

A preferred method for generating loss-of-function phenotypes is by double-stranded RNA interference (dsRNAi), which has been shown to be very effective in both *C. elegans* (Fire *et al.*, Nature (1998) 391:806-811) and *Drosophila* (Kennerdall and

Carthew, Cell (1998) 95:1017-1026). Briefly, complementary sense and antisense RNAs derived from a substantial portion of the gene of interest are synthesized *in vitro*. Phagemid DNA templates containing cDNA clones of the gene are inserted between opposing promoters for T3 and T7 phage RNA polymerases. Alternatively, PCR products can be amplified from coding regions of the gene of interest, where the primers used for the PCR reactions are modified by the addition of phage T3 and T7 promoters. The resulting sense and antisense RNAs are annealed in an injection buffer. In another embodiment, the interfering double-stranded RNA can be generated *in vivo* by co-expression of the complimentary sense and antisense RNAs derived from the gene of interest in the same cells. Interfering double-stranded RNA is administered to the animals usually by injection or by soaking the animals in a solution containing the double-stranded RNA. The animals and their progeny are then inspected for phenotypes of interest.

Peptide and RNA Aptamers

Another method for generating loss-of-function phenotypes is the use of peptide aptamers, which are peptides or small polypeptides that act as dominant inhibitors of protein function. Peptide aptamers specifically bind to target proteins, blocking their function (Kolonin and Finley, Proc. Natl. Acad. Sci. (1998) 95:14266-71). Due to the highly selective nature of peptide aptamers, they may be used not only to target a specific protein, but also to target specific functions of a given protein (*e.g.* a DNA binding function). Further, peptide aptamers may be expressed in a controlled fashion by use of promoters which regulate expression in a temporal, spatial or inducible manner. Peptide aptamers act dominantly; therefore, they can be used to analyze proteins for which loss-of-function mutants are not available.

Peptide aptamers that bind with high affinity and specificity to a target protein may be isolated by a variety of techniques known in the art. In one method, they are isolated from random peptide libraries by yeast two-hybrid screens (Xu *et al.*, Proc. Natl. Acad. Sci. (1997) 94:12473-78). They can also be isolated from phage libraries (Hoogenboom *et al.*, Immunotechnology (1998) 4:1-20).

RNA aptamers are specific RNA ligands for proteins, that can specifically inhibit protein function of the gene (Good *et al.*, Gene Therapy (1997) 4:45-54; Ellington. *et al.*, Biotechnol. Annu. Rev. (1995) 1:185-214). *In vitro* selection methods can be used to identify RNA aptamers having a selected specificity (Bell *et al.*, J. Biol. Chem. (1998) 273:14309-14). RNA aptamers can be used to decrease the expression of an SREBP pathway protein or derivative thereof, or a protein that interacts with any one of SREBP, S1P, S2P, and SCAP.

Transgenesis

Methods are well known for incorporating exogenous nucleic acid sequences into the genome of animals or cultured cells to create transgenic animals or recombinant cell lines. For invertebrate animal models, the most common methods involve the use of transposable elements. There are several suitable transposable elements that can be used to incorporate nucleic acid sequences into the genome of model organisms. Transposable elements are particularly useful for inserting sequences into a gene of interest so that the encoded protein is not properly expressed, creating a "knock-out" animal having a loss-of-function phenotype. Techniques are well-established for the use of P element in *Drosophila* (Rubin and Spradling, Science (1982) 218:348-53; U.S. Pat. No. 4,670,388) and Tc1 in *C. elegans* (Zwaal *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1993) 90:7431-7435; and *Caenorhabditis elegans*: Modern Biological Analysis of an Organism (1995) Epstein and Shakes, Eds.). Other Tc1-like transposable elements can be used such as "minos" (U.S. Pat. No. 5,348,874), "mariner" (Robertson, Insect Physiol. (1995) 41:99-105), and "sleeping beauty"(Ivics *et al.*, Cell (1997) 91(4):501-510). Additionally, several transposable elements, that appear to function in a variety of diverse species, have been identified, including "piggyBac" (Thibault *et al.*, Insect Mol Biol (1999) 8(1):119-23), "hobo" (Atkinson *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1993) 90:9693-9697), and "hermes" (Sarkar *et al.*, Insect Biochem & Molec. Biol. (1997) 27:359-363).

P elements, or marked P elements, are preferred for the isolation of loss-of-function mutations in *Drosophila* SREBP pathway genes because of the precise molecular mapping of these genes, depending on the availability and proximity of preexisting P element insertions for use as a localized transposon source (Hamilton and Zinn, Methods in Cell Biology (1994) 44:81-94; and Wolfner and Goldberg, Methods in Cell Biology (1994) 44:33-80). Typically, modified P elements are used which contain one or more elements that allow detection of animals containing the P element. Most often, marker genes are used that affect the eye color of *Drosophila*, such as derivatives of the *Drosophila white* or *rosy* genes (Rubin and Spradling, Science (1982) 218(4570):348-353; and Klemenz *et al.*, Nucleic Acids Res. (1987) 15(10):3947-3959). However, in principle, any gene can be used as a marker that causes a reliable and easily scored phenotypic change in transgenic animals. Various other markers include bacterial plasmid sequences having selectable markers such as ampicillin resistance (Steller and Pirrotta, EMBO. J. (1985) 4:167-171); and *lacZ* sequences fused to a weak general promoter to detect the presence of enhancers with a developmental expression pattern of interest (Bellen *et al.*, Genes Dev. (1989) 3(9):1288-1300). Other examples of marked P elements useful for mutagenesis have been reported (Nucleic Acids Research (1998) 26:85-88; and http://flybase.bio.indiana.edu).

A preferred method of transposon mutagenesis in *Drosophila* employs the "local hopping" method described by Tower *et al.* (Genetics (1993) 133:347-359). Each new P insertion line can be tested molecularly for transposition of the P element into the SREBP pathway gene of interest by assays based on PCR. For each reaction, one PCR primer is used that is homologous to sequences contained within the P element and a second primer is homologous to the coding region or flanking regions of the SREBP pathway gene. Products of the PCR reactions are detected by agarose gel electrophoresis. The sizes of the resulting DNA fragments reveal the site of P element insertion relative to the SREBP pathway gene. Alternatively, Southern blotting and restriction mapping using DNA probes derived from genomic DNA or cDNAs of the SREBP pathway gene can be used to detect transposition events that rearrange the genomic DNA of the gene. P transposition events that map to the SREBP pathway gene can be assessed for phenotypic effects in heterozygous or homozygous mutant *Drosophila*.

In another embodiment, *Drosophila* lines carrying P insertions in an SREBP pathway gene, can be used to generate localized deletions using known methods (Kaiser, Bioassays (1990) 12(6):297-301; Harnessing the power of *Drosophila* genetics, In Drosophila melanogaster: Practical Uses in Cell and Molecular Biology, Goldstein and Fyrberg, Eds., Academic Press, Inc. San Diego, California). This is particularly useful if no P element transpositions are found that disrupt a particular SREBP pathway gene of interest. Briefly, flies containing P elements inserted near an SREBP pathway gene are exposed to a further round of transposase to induce excision of the element. Progeny in which the transposon has excised are typically identified by loss of the eye color marker associated with the transposable element. The resulting progeny will include flies with either precise or imprecise excision of the P element, where the imprecise excision events often result in deletion of genomic DNA neighboring the site of P insertion. Such progeny are screened by molecular techniques to identify deletion events that remove genomic sequence from the gene of interest, and assessed for phenotypic effects in heterozygous and homozygous mutant *Drosophila*.

In *C. elegans*, Tc1 transposable element can be used for directed mutagenesis of a gene of interest. Typically, a Tc1 library is prepared by the methods of Zwaal *et al.*, *supra* and Plasterk, *supra*, using a strain in which the Tc1 transposable element is highly mobile and present in a high copy number. The library is screened for Tc1 insertions in the region of interest using PCR with one set of primers specific for Tc1 sequence and one set of gene-specific primers. As described in detail in Example 4 below, using such procedures, *C. elegans* strains have been isolated that contain Tc1 transposon insertions within the

SREBP gene. The screen for Tc1 deletions is performed and deletion animals are recovered.

In addition to creating loss-of-function phenotypes, transposable elements can be used to incorporate an SREBP pathway gene, or mutant or derivative thereof, as an additional gene into any region of an animal's genome resulting in mis-expression (including over-expression) of the gene. Alternatively, homologous recombination or gene targeting techniques can be used to substitute the gene for one or both copies of the animal's homologous gene. The transgene can be under the regulation of either an exogenous or an endogenous promoter element, and be inserted as either a minigene or a large genomic fragment. In one application, gene function can be analyzed by ectopic expression, using, for example, *Drosophila* (Brand *et al.*, Methods in Cell Biology (1994) 44:635- 654) or *C. elegans* (Mello and Fire, Methods in Cell Biology (1995) 48:451-482).

Typically, transgenic animals are created that contain gene fusions of the coding regions of the SREBP pathway gene (from either genomic DNA or cDNA) operably joined to a specific promoter and transcriptional enhancer whose regulation has been well characterized, preferably heterologous promoters/enhancers (i.e. promoters/enhancers that are non-native to the SREBP pathway genes being expressed). Heat shock promoters/enhancers, useful for temperature induced mis-expression in *Drosophila* include the *hsp70* and *hsp83* genes, and in *C. elegans*, include *hsp 16-2* and *hsp 16-41*. Tissue specific promoters/enhancers are also useful, and in *Drosophila*, include *sevenless* (Bowtell *et al.*, Genes Dev. (1988) 2(6):620-34), *eyeless* (Bowtell *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1991) 88(15):6853-7), and *glass*-responsive promoters/enhancers (Quiring *et al.*, Science (1994) 265:785-9) which are useful for expression in the eye; and enhancers/promoters derived from the *dpp* or *vestigal* genes which are useful for expression in the wing (Staehling-Hampton *et al.*, Cell Growth Differ. (1994) 5(6):585-93; Kim *et al.*, Nature (1996) 382:133-138). Finally, where it is necessary to restrict the activity of dominant active or dominant negative transgenes to regions where the pathway is normally active, it may be useful to use endogenous promoters of genes in the pathway, such as the SREBP pathway genes.

In *C. elegans*, examples of useful tissue specific promoters/enhancers include the *myo-2* gene promoter, useful for pharyngeal muscle-specific expression; the *hlh-1* gene promoter, useful for body- muscle-specific expression; and the *mec-3* gene promoter, useful for touch-neuron-specific gene expression. In a preferred embodiment, gene fusions for directing the mis-expression of SREBP pathway genes are incorporated into a transformation vector which is injected into nematodes along with a plasmid containing a dominant selectable marker, such as *rol-6*. Transgenic animals are identified as those

exhibiting a roller phenotype, and the transgenic animals are inspected for additional phenotypes of interest created by mis-expression of the SREBP pathway gene.

In *Drosophila*, binary control systems that employ exogenous DNA regulatory elements and exogenous transcriptional activator proteins, are particularly useful for testing the mis-expression of genes in a wide variety of developmental stage-specific and tissue-specific patterns. Two examples of binary exogenous regulatory systems include the UAS/GAL4 system from yeast (Hay *et al.*, Proc. Natl. Acad. Sci. U.S.A. (1997) 94(10):5195-200; Ellis *et al.*, Development (1993) 119(3):855-65), and the "Tet system" derived from *E. coli* (Bello *et al.*, Development (1998) 125:2193-2202). The UAS/GAL4 system is a well-established and powerful method of mis-expression in *Drosophila* which employs the UAS_G upstream regulatory sequence for control of promoters by the yeast GAL4 transcriptional activator protein (Brand and Perrimon, Development (1993) 118(2):401-15). In this approach, transgenic *Drosophila*, termed "target" lines, are generated where the gene of interest to be mis-expressed is operably fused to an appropriate promoter controlled by UAS_G. Other transgenic *Drosophila* strains, termed "driver" lines, are generated where the GAL4 coding region is operably fused to promoters/enhancers that direct the expression of the GAL4 activator protein in specific tissues, such as the eye, wing, nervous system, gut, or musculature. The gene of interest is not expressed in the target lines for lack of a transcriptional activator to drive transcription from the promoter joined to the gene of interest. However, when the UAS-target line is crossed with a GAL4 driver line, mis-expression of the gene of interest is induced in resulting progeny in a specific pattern that is characteristic for that GAL4 line. The technical simplicity of this approach makes it possible to sample the effects of directed mis-expression of the gene of interest in a wide variety of tissues by generating one transgenic target line with the gene of interest, and crossing that target line with a panel of pre-existing driver lines.

In the "Tet" binary control system, transgenic *Drosophila* driver lines are generated where the coding region for a tetracycline-controlled transcriptional activator (tTA) is operably fused to promoters/enhancers that direct the expression of tTA in a tissue-specific and/or developmental stage-specific manner. The driver lines are crossed with transgenic *Drosophila* target lines where the coding region for the gene of interest to be mis-expressed is operably fused to a promoter that possesses a tTA-responsive regulatory element. When the resulting progeny are supplied with food supplemented with a sufficient amount of tetracycline, expression of the gene of interest is blocked. Expression of the gene of interest can be induced at will simply by removal of tetracycline from the food. Also, the level of expression of the gene of interest can be adjusted by varying the level of tetracycline in the food. Thus, the use of the Tet system as a binary control mechanism for mis-expression has

the advantage of providing a means to control the amplitude and timing of mis-expression of the gene of interest, in addition to spatial control. Consequently, if a gene of interest (*e.g.* a tumor suppressor gene) has lethal or deleterious effects when mis-expressed at an early stage in development, such as the embryonic or larval stages, the function of the gene of interest in the adult can still be assessed by adding tetracycline to the food during early stages of development and removing tetracycline later so as to induce mis-expression only at the adult stage.

Dominant negative mutations, where a mutation to a gene creates an inactive protein, can result in loss-of-function or reduced-function phenotype even in the presence of a normal copy of the gene, can be made using known methods (Hershkowitz, Nature (1987) 329:219-222). In the case of active monomeric proteins, over expression of an inactive form, achieved for example, by linking the mutant gene to a highly active promoter, can cause competition for natural substrates or ligands sufficient to significantly reduce net activity of the normal protein. Alternatively, changes to active site residues can be made to create a virtually irreversible association with a target.

In the case of active multimeric proteins, several strategies can guide selection of a dominant negative mutant. In one embodiment, activity of a multmeric complex can be decreased by expression of genes coding exogenous protein fragments that bind to the association domains of the wild type proteins and prevent multimer formation. Alternatively, over-expression of an inactive protein unit can sequester wild-type active units in inactive multimers, and thereby decrease multimeric activity (Nocka *et al.*, EMBO J. (1990) 9:1805-1813). For example, in the case of multimeric DNA binding proteins, the DNA binding domain can be deleted, or the activation domain deleted. Also, in this case, the DNA binding domain unit can be expressed without the activation domain causing sequestering of the target DNA. Thereby, DNA binding sites are tied up without any possible activation of expression. In the case where a particular type of unit normally undergoes a conformational change during activity, expression of a rigid unit can also inactivate resultant complexes. It is also possible to replace an activation domain with a transcriptional repression domain and thus change a transcriptional activator into a transcriptional repressor. Transcriptional repression domains from the engrailed and Kruppel proteins have been used for such a purpose (Jaynes and O'Ferrell, EMBO J. (1991) 10:1427-1433; Licht. *et al.*, Proc. Natl. Acad. Sci. USA (1993) 90:11361-65).

Expression Analysis of SREBP pathway genes

Various expression analysis techniques may be used to identify genes which are differentially expressed between a cell line or an animal expressing a wild type SREBP

pathway gene compared to another cell line or animal expressing a mutant SREBP pathway gene. Such expression profiling techniques include differential display, serial analysis of gene expression (SAGE), nucleic acid array technology, subtractive hybridization, and proteome analysis (*e.g.* mass-spectrometry and two-dimensional protein gels). Nucleic acid array technology may be used to determine a global (i.e., genome-wide) gene expression pattern in a normal animal for comparison with an animal having a mutation in one or more SREBP pathway gene. Gene expression profiling can also be used to identify other genes (or proteins) that may have a functional relation to (*e.g.* may participate in a signaling pathway with) or be a transcriptional target of an SREBP pathway gene. The genes are identified by detecting changes in their expression levels following mutation, i.e., insertion, deletion or substitution in, or over-expression, under-expression, mis-expression or knock-out, of an SREBP pathway gene.

Phenotypes associated with SREBP pathway gene mutations

After isolation of model animals carrying mutated or mis-expressed SREBP pathway genes or inhibitory RNAs, animals are carefully examined for phenotypes of interest. For analysis of SREBP pathway genes that have been mutated (i.e. deletions, insertions, and/or point mutations) animal models that are both homozygous and heterozygous for the altered SREBP pathway gene are analyzed. Examples of specific phenotypes that may be investigated include lethality; sterility; and changes in various characteristics of the animal such as motility, body shape, body size and weight, metabolism, lipid accumulation, feeding, development, morphogenesis of organs, brood size, thermotaxis, etc. Some phenotypes more specific to flies include alterations in: morphogenesis of the peripheral sensory organs, imaginal discs, eye development, wing development, leg development, bristle development, antennae development, gut development, fat body, and musculature. Some phenotypes more specific to nematodes include: alterations in chemotaxis, a dauer constitutive phenotype, a dauer defective phenotype, and a pale-intestine phenotype. A phenotype of particular interest in *C. elegans* is the pale intestine phenotype, which is indicative of defects in lipid metabolism and is discussed in more detail below and in the Examples.

Genomic sequences containing an SREBP pathway gene can be used to confirm whether an existing mutant *Drosophila* or *C. elegans* line corresponds to a mutation in one or more SREBP pathway genes, by rescuing the mutant phenotype. Briefly, a genomic fragment containing the SREBP pathway gene of interest and potential flanking regulatory regions can be subcloned into any appropriate *Drosophila* or *C. elegans* transformation vector, and injected into the animals. For *Drosophila*, an appropriate helper plasmid is used

in the injections to supply transposase. Resulting transformants are crossed for complementation testing to an existing panel of *Drosophila* or *C. elegans* lines whose mutations have been mapped to the vicinity of the gene of interest (Fly Pushing: The Theory and Practice of Drosophila Genetics, (1997) Cold Spring Harbor Press, Plainview, NY; and *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, *supra*. If a mutant line is discovered to be rescued by this genomic fragment, as judged by complementation of the mutant phenotype, then the mutant line likely harbors a mutation in the SREBP pathway gene. This prediction can be further confirmed by sequencing the SREBP pathway gene from the mutant line to identify the lesion in the SREBP pathway gene.

Identification of genes that modify SREBP pathway genes

The characterization of new phenotypes created by mutations in SREBP pathway genes enables one to test for genetic interactions between SREBP pathway genes and other genes that may participate in the same, related, or interacting genetic or biochemical pathway(s). Individual genes can be used as starting points in large-scale genetic modifier screens as described in more detail below. Alternatively, RNAi methods can be used to simulate loss-of-function mutations in the genes being analyzed. It is of particular interest to investigate whether there are any interactions of SREBP pathway genes with other well-characterized genes, particularly genes involved in lipid metabolism. For example, a candidate gene that may be tested for interaction with the SREBP pathway is the insulin receptor gene (referred to as *inr* in *Drosophila*, and *daf-2* in *C. elegans*).

Genetic Modifier Screens

A genetic modifier screen using invertebrate model organisms is a particularly preferred method for identifying genes that interact with SREBP pathway genes, because large numbers of animals can be systematically screened making it more likely that interacting genes will be identified. In *C. elegans* and *Drosophila*, a screen of up to about 10,000 animals is considered to be a pilot-scale screen. Moderate-scale screens usually employ about 10,000 to about 50,000 flies or up to about 100,000 worms, and large-scale screens employ greater than about 50,000 or 100,000 flies or worms, respectively. In a genetic modifier screen, animals having a mutant phenotype due to a mutation in one or more SREBP pathway genes are further mutagenized, for example by chemical mutagenesis or transposon mutagenesis. The mutagenesis procedures used in typical genetic modifier screens of *C. elegans* are well known in the art. One method involves exposure of hermaphrodites that carry mutations in one or more SREBP pathway genes to a mutagen,

such as EMS or trimethylpsoralen with ultraviolet radiation (Huang and Sternberg, Methods in Cell Biology (1995) 48:97-122). Alternatively, transposable elements are used, oftentimes by the introduction of a mutator locus, such as *mut-2*, which promotes mobility of transposons (Anderson, Methods in Cell Biology (1995) 4:31-58).

In *Drosophila*, the mutagenesis methods and other procedures used in a genetic modifier screen depend upon the precise nature of the mutant allele being modified; these methods are discussed in more detail below under the *Drosophila* genetic modifier screen subheading.

Progeny of the mutagenized animals are generated and screened for the rare individuals that display suppressed or enhanced versions of the original mutant SREBP pathway phenotype. Such animals are presumed to have mutations in other genes, called "modifier" genes, that participate in the same phenotype-generating pathway. The newly-identified modifier genes can be isolated away from the mutations in the SREBP pathway genes by genetic crosses, so that the intrinsic phenotypes caused by the modifier mutations can be assessed in isolation.

Modifier genes can be mapped using a combination of genetic and molecular methods known in the art. Modifiers that come from a genetic screen in *C. elegans* are preferably mapped with visible genetic markers and/or with molecular markers such as STS markers (The Nematode *Caenorhabditis elegans, supra*; *Caenorhabditis elegans*: Modern Biological Analysis of an Organism, *supra*). Modifier genes may be uncovered by identification of a genomic clone which rescues the mutant phenotype, as described above. Alternatively, modifier genes that are identified by a Tc1-based screen can be uncovered using transposon display technology (Korswagen *et al.*, Proc Natl Acad Sci U.S.A. (1996) 93(25):14680-5).

Standard techniques used for the mapping of modifiers that come from a genetic screen in *Drosophila* include meiotic mapping with visible or molecular genetic markers; complementation analysis with deficiencies, duplications, and lethal P-element insertions; and cytological analysis of chromosomal aberrations (Fly Pushing: Theory and Practice of Drosophila Genetics, *supra*; *Drosophila*: A Laboratory Handbook, *supra*). Genes corresponding to modifier mutations that fail to complement a lethal P-element may be cloned by plasmid resuce of the genomic sequence surrounding that P-element. Alternatively, modifier genes may be mapped by phenotype rescue and positional cloning (Sambrook *et al.*, *supra*).

Newly identified modifier mutations can be tested directly for interaction with other genes of interest known to be involved or implicated in the SREBP using methods described above. Also, the new modifier mutations can be tested for interactions with genes

in other pathways that are not believed to be related to SREBP signaling (*e.g. Notch* in *Drosophila*, and *lin* in *C. elegans*). New modifier mutations that exhibit specific genetic interactions with other genes implicated in lipid metabolism, but not interactions with genes in unrelated pathways, are of particular interest.

The modifier mutations may also be used to identify "complementation groups". Two modifier mutations are considered to fall within the same complementation group if animals carrying both mutations in *trans* exhibit essentially the same phenotype as animals that are homozygous for each mutation individually and, generally, are lethal when in trans to each other (Fly Pushing: The Theory and Practice of Drosophila Genetics, *supra*). Generally, individual complementation groups defined in this way correspond to individual genes.

When SREBP pathway modifier genes are identified, homologous genes in other species can be isolated using procedures based on cross-hybridization with modifier gene DNA probes, PCR-based strategies with primer sequences derived from the modifier genes, and/or computer searches of sequence databases. For therapeutic applications related to the function of SREBP pathway, human and rodent homologues of the modifier genes are of particular interest. For pesticide and other agricultural applications, homologues of modifier genes in insects and arachnids are of particular interest. Insects, arachnids, and other organisms of interest include, among others, Isopoda; Diplopoda; Chilopoda; Symphyla; Thysanura; Collembola; Orthoptera, such as *Blattella germanica*; Dermaptera; Isoptera; Anoplura; Mallophaga; Thysanoptera; Heteroptera; Homoptera, including *Bemisia tabaci*, and *Myzus* spp.; Lepidoptera including *Plodia interpunctella, Pectinophora gossypiella*, *Plutella* spp., *Heliothis* spp., and *Spodoptera* species; Coleoptera such as *Leptinotarsa decemlineata*, *Diabrotica* spp., *Anthonomus* spp., and *Tribolium* spp.; Hymenoptera, including *Apis mellifera*; Diptera, including *Anopheles* spp.; Siphonaptera, including *Ctenocephalides felis*; Arachnida; and Acarinan, including *Amblyoma americanum*; and nematodes, including *Meloidogyne* spp., and *Heterodera glycinii.*

Genetic modifier screens in Drosophila

The procedures involved in typical *Drosophila* genetic modifier screens are well-known in the art (Wolfner and Goldberg, Methods in Cell Biology (1994) 44:33-80; and Karim *et al.*, Genetics (1996) 143:315-329). The procedures used differ depending upon the precise nature of the mutant allele being modified. If the mutant allele is genetically recessive, as is commonly the situation for a loss-of-function allele, then most typically males, or in some cases females, which carry one copy of the mutant allele are exposed to an effective mutagen, such as EMS, MMS, ENU, triethylamine, diepoxyalkanes, ICR-170,

formaldehyde, X-rays, gamma rays, or ultraviolet radiation. The mutagenized animals are crossed to animals of the opposite sex that also carry the mutant allele to be modified. In the case where the mutant allele being modified is genetically dominant, as is commonly the situation for ectopically expressed genes, wild type males are mutagenized and crossed to females carrying the mutant allele to be modified.

The progeny of the mutagenized and crossed flies that exhibit either enhancement or suppression of the original phenotype are immediately crossed to adults containing balancer chromosomes and used as founders of a stable genetic line. In addition, progeny of the founder adult are retested under the original screening conditions to ensure stability and reproducibility of the phenotype. Additional secondary screens may be employed, as appropriate, to confirm the suitability of each new modifier mutant line for further analysis.

Although the above-described *Drosophila* genetic modifier screens are quite powerful and sensitive, some genes that participate in the SREBP pathway may be missed in this approach, particularly if there is functional redundancy of those genes. This is because the vast majority of the mutations generated in the standard mutagenesis methods will be loss-of-function mutations, whereas gain-of-function mutations that could reveal genes with functional redundancy will be relatively rare. Another method of genetic screening in *Drosophila* has been developed that focuses specifically on systematic gain-of-function genetic screens (Rorth *et al.*, Development (1998) 125:1049-1057). This method is based on a modular mis-expression system utilizing components of the GAL4/UAS system (described above) where a modified P element, termed an "enhanced P" (EP) element, is genetically engineered to contain a GAL4-responsive UAS element and promoter. The resulting transposon is used to randomly tag genes by insertional mutagenesis (similar to the method of P element mutagenesis described above). Thousands of transgenic *Drosophila* strains, termed EP lines, can be generated, each containing a specific UAS-tagged gene. This approach takes advantage of the preference of P elements to insert at the 5'-ends of genes. Consequently, many of the genes that are tagged by insertion of EP elements become operably fused to a GAL4-regulated promoter, and increased expression or mis-expression of the randomly tagged gene can be induced by crossing in a GAL4 driver gene.

Systematic gain-of-function genetic screens for modifiers of phenotypes induced by mutation or mis-expression of an SREBP pathway gene can be performed by crossing several thousand *Drosophila* EP lines into a genetic background containing a mutant or mis-expressed SREBP pathway gene, and further containing an appropriate GAL4 driver transgene. The progeny of this cross are then analyzed for enhancement or suppression of the original mutant phenotype as described above. Those identified as having mutations

that interact with the SREBP pathway can be crossed further to verify the reproducibility and specificity of this genetic interaction. EP insertions that demonstrate a specific genetic interaction with a mutant or mis-expressed SREBP pathway gene, have a physically tagged a new gene which can be identified and sequenced using PCR or hybridization screening methods, allowing the isolation of the genomic DNA adjacent to the position of the EP element insertion.

BODIPY-fatty acid conjugates for determining lipid content of nematodes

Because defects in the SREBP pathway can result in abnormal metabolism of lipids, a method for readily identifying mutant model organisms that exhibit abnormalities in lipid metabolism would be beneficial. Prior methods for assessing lipid content in nematodes includes the use of non-vital stains such as Sudan Black (Kimura *et al.*, Science (1997) 277:942-6).

However, the drawbacks of these techniques are that the nematodes must be fixed prior to staining. Fixation can introduce artifacts, making an accurate assessment difficult, and furthermore, kills the animals making it impossible to carry out further genetic analysis on the animals that are tested. In order to avoid these problems associated with fixing nematodes, certain vital stains were tried that are routinely used for staining lipid in cultured cells such as Nile Red (Greenspan *et al.*, J Cell Biol, (1985) 100:965- 973). However, it was found that these dyes tended to result in background fluorescence of gut granules which are auto-fluorescent organelles of the intestinal epithelial cells that are thought be to lysosomes. In many cases, these fluorescent vital stains appeared to be concentrated in gut granules, enhancing their fluorescence and causing difficulty in accurately measuring the fluorescence due to lipid droplet staining in the intestine. Accordingly, the invention provides an improved method for measuring lipid storage in live nematodes. It has been found that BODIPY® dyes conjugated to fatty acids (*e.g.* BODIPY® FL C12 (4,4-difluoro-5,7-dimethyl- 4-bora-3a,4a-diaza-s-indacene- 3-dodecanoic acid), and C1-BODIPY® 500/510 C12 (4,4-difluoro-5-methyl-4-bora- 3a,4a-diaza-s-indacene-3-dodecanoic acid) Molecular Probes, Eugene, OR) concentrate in lipid droplets in the intestines of living nematodes. These dyes do not have the drawbacks associated with other vital dyes because, in addition to clearly staining and fluorescing in lipid droplets in the intestine, they quench the background fluorescence due to the gut granules. Accordingly, the invention provides a method of using BODIPY®-fatty acid conjugates to stain live nematodes for determining the relative and absolute lipid content in response to changes in metabolic conditions brought on by a) changes in genetic backgrounds including mutations in genes essential for control of metabolic processes, b) changes in environmental

conditions such as food sources, temperature, and crowding conditions, and c) different developmental states including the dauer larva. This method is particularly valuable in uses that involve genetic screens and compound screens based on changes in metabolic processes such as the SREBP processing pathway, among others. The method allows considerable increases in accuracy of lipid quantification *in vivo* over the use of other fluorescent lipophilic stains, making automated sorting of the nematodes based on fluorescence feasible.

BODIPY® conjugates have previously been used to study (1) lipid content in the surface membrane of *Shistosoma mansoni* worms (Redman and Kusel, Parasitology (1996) 113(2):137- 143), (2) lipid endocytosis in cultured mammalian fibroblasts (Pagano and Chen, Ann N Y Acad Sci (1998) 845:152-160), (3) lipid trafficking between the Golgi apparatus and plasma membrane of cultured mammalian fibroblasts (Pagano *et al.*, J. Cell. Biol (1991) 113(6):1267-1279), (4) fatty acid transport by *Saccharomyces* (Faergeman *et al.*, J. Biol. Chem (1997) 272(13):8531-8538) and (5) distribution of ivermectin in muscle vesicle membranes of *Ascaris suum* (Marin and Kusel, Parasitology (1992) 104(3):549-555). However, these prior uses of BODIPY® conjugates do not suggest the applicability of BODIPY® conjugates, and in particular, BODIPY® fatty acid conjugates, for quantification of lipid storage in nematodes. Moreover, the fact that BODIPY® fatty acid conjugates quenches background fluorescence from lysosomes, providing for more accurate quantification, is an unexpected and important advantage provided by the invention that permits large-scale, automated sorting of animals based on fluorescence.

BODIPY®-fatty acid conjugates can be used to stain nematodes of different genetic backgrounds for use in genetic screens, both *de novo* screens for mutations affecting lipid content of whole nematodes and modifier screens for mutations that change lipid accumulation in mutant nematodes (for example, the insulin receptor (*daf-2*) or the SREBP homolog (*pin-1)* nematodes). The intestines of the nematodes can be visually examined for lipid content under a fluorescent microscope and mutant animals can be subsequently propagated for cloning purposes. This method can be used in conjunction with automatic flow sorter technology to rapidly separate large numbers of living nematodes by lipid content. This would be useful either for automated high throughput genetic screening or for large scale automated separation of dauer larvae from other developmental stages. Additionally, the method can be used to determine changes in lipid accumulation in nematodes exposed to inhibitory compounds that might serve as therapeutic agents for the control of diabetes, obesity, lipid storage diseases, or other human or animal diseases. A test compound can be administered to a nematode by direct contact, ingestion, injection, or any suitable method and changes in lipid content of the nematode or its progeny are

observed. Further, the method is applicable to reverse genetic screening using inhibitory RNA. For example, nematodes could be exposed to combinations of large numbers of RNAs in 384-well plates and screened for changes in lipid content mediated by RNAi using fluorometry or direct visual observation.

EXAMPLES

The following examples show how the nucleic acid sequences of SEQ ID NOs 1, 3, and 5 were isolated, and how these sequences, and derivatives and fragments thereof, as well as other SREBP pathway nucleic acids and gene products can be used for genetic studies to elucidate mechanisms of the SREBP pathway as well as the discovery of potential pharmaceutical or pesticidal agents that interact with the pathway. As used herein, all *C. elegans*-derived gene sequences are designated by the letters "ce" in front of the gene sequence. Likewise, all *Drosophila*-derived gene sequences are designated by the letter "d" in front of the gene sequence.

These Examples are provided merely as illustrative of various aspects of the invention and should not be construed to limit the invention in any way.

EXAMPLE 1: CLONING OF *C. ELEGANS* SREBP

The *C. elegans* genomic database was searched with the protein sequence of the human SREBP-1, SREBP-2, and *Drosophila melanogaster* SREBP homologue, HLH106, using the TBLASTN search tool (Altschul *et al.*, *supra*). One *C. elegans* open reading frame showed significant homology with all three of the above SREBP proteins. This homology extends throughout most of the SREBP protein sequences. The *C. elegans* open reading frame is located on two overlapping clones on the right arm of chromosome III (Y47D3 and H10N23). At the time of the search, there were no previous annotations, gene predictions, or candidate mutants that mapped to this region that would suggest previous identification of this open reading frame as an SREBP-related gene.

Using BLAST analyses (Altschul *et al.*, *supra*) and the GENSCAN Genefinder program (Burge and Karlin, J. Mol. Biol. (1997) 268(1):78-94) , a predicted exon-intron structure for the *C. elegans* SREBP-related gene (ceSREBP) was generated. This *C. elegans* homologue of SREBP cDNA was cloned in order to validate its existence as an expressed mRNA, and to determine the cDNA and protein sequence for the elucidation of ceSREBP function. Moreover, cloning of ceSREBP is a prerequisite for future genetic manipulations that require knowledge of the sequence, such as RNAi experiments, generation of misexpression constructs, isolation of Tc1 insertion or chemical deletion mutants, etc.

Cloning strategy:

The N-terminal and C-terminal ends of ceSREBP were cloned using gene-specific internal primers and non-specific primers at 3' and 5' ends. Internal primers were made to regions of high homology according to the GENSCAN prediction for ceSREBP, that were also predicted by the ACEDB Genefinder (Richard Durbin and Jean Thierry Mieg (1991-present), A *C. elegans* Database. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov). They were designed to amplify the ends of the cDNA, not the full-length cDNA. Once the end sequence was known, the full-length cDNA was amplified in overlapping N-terminal and C-terminal parts using gene-specific primer pairs. The template for amplification was a mixed-stage, 1st strand cDNA pool that was synthesized from poly-A+ RNA using the NotI primer/adapter (Life Technologies, Gaithersburg, MI).

N-terminal:

Antisense, internal primer for amplification of the N-terminal was made to sequence encoding the ceSREBP bHLHz region: GTACGACGCTCGGTTTTTGGTC (SEQ ID NO:1). Sense primer no. 1 was the 5' splice leader (SL1) sequence: GGTTTAATTACCCAAGTTTGAG (SEQ ID NO:1). Amplifications were performed using Expand™ buffers and enzyme mixes (Roche, Summerville, N.J.). Amplification conditions were as follows:

2 µl *C. elegans* (ce) mixed-stage 1st strand cDNA
5 µl 2 mM dNTPs
5 µl Expand™ High Fidelity 10X buffer with $MgCl_2$
3 µl SL1, 5 µM
3 µl ceSREBP, 5 µM
0.75 µl Expand™ High Fidelity enzyme mix
31.25 µl H_2O

94° C 2min
10 cycles of: 94°C 1 sec
52°C 30 sec
68°C 4 min
25 cycles of: 94°C 15 sec
52° C 30 sec
68°C 4 min + 20 sec/cycle
72°C 5min
4° C hold
1 µl Amplitaq™ (Perkin-Elmer, Foster City, CA) added with an additional incubation for 10min at 72°C

PCR products were run on a 1% agarose gel. The major product was a band of about 1kb in size.

C-terminal:

Sense primer no. 2, was an internal primer made to a region of homology between human SREBP1 and Genscan-predicted ceSREBP, encoding LCAVNLAE: CCTCTGTGCAGTAAACCTTGCTG (SEQ ID NO:1). The non-specific antisense primer, pN1, was made to the 5' end of the NotI primer/adapter: GACTACTTCTAGATGGCGAGC (SEQ ID NO:9). The same amplification conditions as for the N-terminal were used using these primers. The product of this amplification was 1.3kb in size.

Cloning and sequencing:

PCR fragments were gel-purified by beta-agarase treatment (NEB, Beverly, MA) cloned into the pCR2.1 vector (Invitrogen, San Diego, CA) essentially according to manufacturer's protocols for ligation and transformation into E. coli. Individual colonies were screened for an insert of the correct size by PCR using M13 forward and reverse primers to pCR2.1. Individual colonies from each transformation were grown up overnight in 3 ml LB-Amp, and DNA was prepared using Easy-Pure Preps™ (super mini) (Primm Labs, Cambridge, MA). DNA preps were digested using EcoRI enzyme to check for clones with correct insert size. Clones were end-sequenced with Big Dye™ dye-terminator sequencing kit (ABI, Foster City, CA) using M13 forward and reverse primers to pCR2.1.

Sequencing conditions were as follows:

1 µl miniprep DNA (~100 ng)
1 ul Big Dye™
1 µl primer, 0.8 µM
0.5 µl 5X buffer (400 mM Tris, pH 9; 10 mM $MgCl_2$)
1.5 µl H_2O.

96°C for 5min
25 cycles of: 96°C for 30 sec
50°C for 15 sec
60°C for 4 min

Reactions were ethanol precipitated and sequenced. Sequencing products were analyzed using Sequencher program (Gene Codes Corporation, Ann Arbor, MI) and BLAST (Altschul *et al.*, *supra*). A single N-term contiguous sequence ("contig") was assembled that shared sequence identity with the YAC sequence Y47D3 (GI:3646936) from which gene predictions were made, and was virtually identical to the gene prediction in this region. Two non-overlapping C-terminal contigs were assembled, one of which contained the sequence of sense primer no.2; the other, pN1. Both shared identity with Y47D3 in BLAST searches to *C. elegans* genomic sequence and showed homology to SREBP

sequences from other species. A likely termination codon and possible poly-A signals were identified.

Amplification of full-length cDNA:

A longer C-terminal fragment was amplified which overlapped with the N-terminal fragment and contained the remainder of the unknown sequence. The ~2.9 kb fragment was amplified using the same conditions as for the N- and C- terminals but with the primers described below. Another difference is that 3 µl ce mixed stage 1st strand cDNA was used rather than 2 µl.. The primers used were a sense primer within the ~1 kb N-terminal fragment, and referred to as Y47-4, and a primer referred to as ceSREBP7, which includes the predicted termination codon:

Y47-4: AGCAATGGAACATATCAACGGG (SEQ ID NO:1)
ceSREBP7: CAATTCAAAGATCCATAGAAGTATG (SEQ ID NO:1)

The ~2.9 kb fragment was cloned into the pCR2.1 vector as described above.

Sequencing the full-length cDNA:

To obtain complete sequence of the full-length gene of ceSREBP, the following set of seven sequencing primers (ceSREBPs1- s7; contained within SEQ ID NO:1) were synthesized based on sequences derived from the original N-terminal and C-terminal contigs, and to the most highly conserved regions of the predicted gene.

sense
ceSREBPs1: GCATGTTCAGCGACGAATGG
ceSREBPs2: GCAACACTACGACGGGCTAT
ceSREBPs3: TGGATTGCTCGCTGGAAGTG
ceSREBPs4: GGAACTTGTCGGTGGTGACG

antisense
ceSREBPs5: CCCTTGAAGCTTTGTGTCCA
ceSREBPs6: CGTGGAAGTCCGTCGTTTGA
ceSREBPs7: GGTCACCATGGATCAGCAGT

These primers, as well as M13 forward and reverse, were used to sequence clones containing the ~1.3 and the ~2.9 kb C-terminal fragments using the above-described sequencing methods. When aligned with previously obtained sequence, these yielded a single open reading frame (ORF) of ~3.4 kb in size which. BLAST analysis against GenBank sequences, showed highest homology to other SREBPs.

Error-free cDNA clones

All the C-terminal clones that contributed to the above contig contained PCR errors (i.e., single nucleotide discrepancies from the other clones). To obtain error-free sequence the ~3 kb C-term fragment was re-amplified using the high-fidelity Pfu enzyme (Stratagene) and, as template, a commercial mixed stage cDNA library (Stratagene), in addition to the 1st strand cDNA pool.

Reaction 1
2 µl library cDNA
5 µl 10X Pfu buffer
10 µl 2 mM dNTPs
10 µl Y47-4, 5 µM
10 µl ceSREBP7, 5 µM
1 µl Pfu
12 µl H_2O

Reaction 2 was the same as the first with the exception that 0.5 µl Amplitaq™ was used, and 0.5 µl Pfu was used instead of 1 µl.

For both reactions:

94°C 2 min
35 cycles of: 94°C 15 sec
55°C 30 sec
72°C 6 min
72°C 10 min
4°C hold

1 µl Amplitaq™ was added with an additional incubation for 10 min at 72°C.

The fragments from both reactions were cloned into the pCRII vector as described above for the pCR2.1 vector. One clone from each of the above reactions was sequenced through using the following primers:

M13 forward and reverse; ceSREBPs1, s2, s3, s4, s6, s7 (above); and the following additional primers ceSREBPs9-s15 (contained within SEQ ID NO:1):

sense
ceSREBPs9: CGTTGGATCGATCGCTTCCA
ceSREBPs10: CCGCCGAAGATTTTGACAGA
ceSREBPs11: TGGGACAAGGGGAGATTGTT

antisense
ceSREBPs12: GAACGTGCGTCCACCATGTG
ceSREBPs13: GCTCCAACCTTTTCGCATCT
ceSREBPs14: GGAGATGATTCGACGGGTGA
ceSREBPs15: TCCCCGGAATCACTATCCTC

These sets of reactions gave full sequence in both directions and gave identical sequence from the two clones.

Two clones were also sequenced that contained the ~1 kb N-terminal fragment with the following additional primers (contained within SEQ ID NO:1):

sense
ceSREBPs8: GCAGCGTCGCTTTTTGTTAA

antisense
ceSREBPs16: TGATGGTGGTGATGAGGTGG
ceSREBPs17: AATTGTTGGGTGGCGGCTAG

These reactions, together with sequence from M13 forward and reverse primers, gave a full sequence in both directions that was nearly identical to the posted, unfinished sequence from Y47D3. The cDNA sequence of the ceSREBP gene, SEQ ID NO:1 is shown in Figure 2. The cDNA is 3419 nucleotides long. This full-length clone contained a single open reading frame with an apparent translational initiation site at nucleotide position 24 and a stop signal at nucleotide position 3365. The predicted polypeptide precursor is 1113 amino acids long. Additional features include an acidic domain at about nucleotides 24 to 233 (amino acid residues 1 to 69); a possible second acidic domain at about nucleotides 987 to 1040 (amino acid residues 321 to 338); a basic Helix-loop-helix domain at about nucleotides 1089 to 1286 (amino acid residues 355 to 421); a first transmembrane domain at about nucleotides 1455 to 1514 (amino acid residues 477 to 497); and a second transmembrane domain at about nucleotides 1653 to 1706 (amino acid residues 543 to 561).

A BLAST analysis against the Y47D3 clone which has a total of 351,956 nucleotides, revealed 12 regions of Y47D3 which share sequence identity with SEQ ID NO:1, as shown in Table II.

TABLE II

Base # of SEQ ID NO:1	Base # of Y47D3	% Sequence Identity
1-80	179,410-179,331	100
81-213	178,918-178,786	100
214-523	178,528-178,218	100
527-632	177,448-177,338	96
633-1052	177,286-176,864	100
1053-1288	176,520-176,285	100
1289-1482	175,768-175,568	100
1483-2011	175,523-174,994	100
2012-2408	174,687-174,288	100
2409-2636	174,228-174,001	100

2637-2790	173,954-173,801	100
2791-3151	155,054-154,694	100
3152-3397	154,638-154,393	100

An alignment of the predicted protein sequence (SEQ ID NO:2) against the human and *Drosophila* SREBP proteins was performed. Amino acid residues 353 to 423 of SEQ ID NO:2 share 45% and sequence identity and 77% sequence similarity with amino acid residues 281-351 of *Drosophila melanogaster* SREBP (SEQ ID NO:8; Theopold *et al.*, *supra*; GI079656). Amino acid residues 466 to 826 of SEQ ID NO:2 share 28% sequence identity and 47% sequence similarity with human SREBP2 (GI1082805). The presence of other gene and protein sequences bearing significant homology to ceSREBP was investigated using the BLAST family of computer programs (Altschul *et al.*, *supra*). The amino acid sequence of a sterol regulatory element-binding protein-1 (SREBP-1) from *Mus musculus* (GI4240012) was most similar, sharing 52% sequence identity and 71% sequence similarity with amino acid residues 335-428 of SEQ ID NO:2 and having up to 5 contiguous identical amino acids in common with SEQ ID NO:2. Sequence similarity, to a lesser extent, was revealed between SEQ ID NO:2 and sequences from U.S. Pat. No. 5,780,262 (GI3998144), and others.

The presence of other gene and protein sequences bearing significant homology to ceSREBP was investigated using the BLAST family of computer programs against public databases. The following amino acid sequences were the most similar: SREBP-1, Chinese Hamster (GI 1083186); SREBP-1, *Cricetulus griseus* (GI 516003); Sequence 54 from patent US 5527690 (GI 1610915); SREBP2 precursor, human (GI 1082805); SREBP-2, *Homo sapiens* (GI 451330); SREBP2 precursor, Chinese hamster (GI 1083185); Sequence 38 from patent US 5527690 (GI 1610908); SREBP-1, *Homo sapiens* (GI 409405); SREBP-2, *Cricetulus griseus* (GI 551506); Transcription factor ADD1, Rat (GI 540006); and HLH106, *Drosophila Melanogaster* (GI 107965).

Subsequent to the above analysis, a Genefinder prediction of the ceSREBP protein was entered into the Genbank database, which is 100% identical to SEQ ID NO:2, and is designated GI 3881008.

EXAMPLE 2: ceSREBP EXPRESSION ANALYSIS

Strategy

Expression of ceSREBP was assayed using a transcriptional reporter system in which the putative promoter/enhancer region of ceSREBP was fused to GFP. To determine how much genomic sequence to include in the reporter construct, the Y47D3 contig

containing the N-terminal region of the ceSREBP cDNA and ~25 kb upstream of the ceSREBP initiation codon, was analyzed using ACEDB Genefinder and GENSCAN programs (Burge and Karlin, *supra*). There were no known genes within this region, and no predicted genes reported by either program. Of the two predicted genes within ~8 kb of ceSREBP, one, ~5 kb upstream of ceSREBP showed limited homologies by BLAST analysis to *C. elegans* expressed sequence tags (EST). A genomic fragment of ~4.5 kb was chosen as the putative promoter/enhancer region.

Amplification of genomic enhancer/promoter region

PCR primers were designed to amplify the ~4.5 kb genomic fragment, including the first few amino acids of ceSREBP. Restriction sites were included in the primers to facilitate subcloning into the GFP reporter vector pPD117.01 (from the laboratory of Dr. Andrew Z. Fire (Fire Lab)), Carnegie Institution of Washington, Baltimore, MD) in an in-frame translational fusion to GFP. The sense primer, nucleotides 71,242-71,265 of Y47D3 (GI:3646936), contained an AscI site; the antisense primer, nucleotides 66,719-66,747 of Y47D3, contained a KpnI/Asp718 site:

ceSREBPp1: ATGGGCGCGCCAACCAAAGTGTGATGCAACAG (SEQ ID NO:28)
ceSREBPp2: GAGGGTACCTCGTTCATTCTGAAAAAAAAAAGTC (SEQ ID NO:29)

Amplification was done in duplicate to provide two independently-amplified promoter fragments for independent confirmation of the expression pattern; conditions were as follows:

2 µl N2 genomic DNA (50 ng/µl)
5 µl 10X Klentaq™ buffer (Clontech, Palo Alto, CA)
1 µl 10 mM dNTPs
5 µl ceSREBPp1, 5 µM
5 µl ceSREBPp2, 5 µM
1 µl Klentaq™ enzyme
31 µl H_2O
94°C 2 min
25 cycles of: 94°C 15 sec
52°C 1 min
72°C 5 min
72°C 10 min
4°C hold

1 µl Amplitaq™ added with an additional incubation for 10 min at 72°C.

One half of each reaction was run on a 1% Seaplaque GTG, 1X TAE gel, and the major product at ~4.5 kb excised. The fragments were purified using Geneclean (Bio101) and subcloned into the pCRII vector essentially according to manufacturer's protocols (Invitrogen) for ligation and transformation. DNA from individual colonies from each

transformation was prepared using Easy-Pure preps (super mini). DNA preps were checked using AscI+Asp718 and AscI+NotI restriction digests. Clones that appeared as expected by digest were end-sequenced with Big Dye™ dye-terminator sequencing kit using M13 forward and reverse primers to pCRII.

Sequences were analyzed using Sequencher. From each original PCR reaction, a single clone that contained the expected insert was identified.

To subclone the promoter fragment into the GFP reporter vector, this vector, and the promoter fragments in pCRII were digested with AscI+Asp718, gel-purified using Geneclean (Bio101), ligated together and transformed into E. coli using standard procedures. DNA from individual colonies from each transformation was prepared using Easy-Pure™ preps (super mini). DNA preps were checked using AscI+Asp718, HincII, ClaI, DraIII, EcoRI, and HindIII restriction digests. Several clones from each original PCR reaction were checked by end-sequencing with primers that sequenced through the two cloning junctions. Colonies of clones that looked correct by sequence analysis were re-streaked. DNA from individual colonies was prepared by Qiagen midi preps and checked by restriction digest.

Expression analysis – ceSREBP::GFP

By GFP expression analysis, ceSREBP is first expressed weakly in embryonic gut cells at the time of gut cell polarization, which marks the beginning of differentiation. There is strong fluorescence by the "bean stage" which persists in all intestinal cells throughout embryogenesis and at all larval and adult stages. There is also weak fluorescence in the pharynx. Because there is high specificity of expression of ceSREBP in intestinal cells, the ceSREBP promoter, contained within nucleotides 66,719-71,265 of Y47D3 (GI:3646936), has utility as a tissue specific promoter that can be operably linked to heterologous sequences, such as marker genes and/or genes of interest. Thus, the ceSREBP promoter can be used for studying biochemical pathways within the intestine of *C. elegans*.

EXAMPLE 3: RNA INTERFERENCE (RNAI) OF *C. ELEGANS* SREBP, S2P and SCAP

Methods:

PCR was carried out on *C. elegans* sequences for SREBP (SEQ ID NO:1) and S2P (Rawson *et al.*, *supra*; GI1559384), and a Genbank sequence (GI3875380), that is annotated as having HMG-CoA reductase homology, and additionally has been determined to have homology to the human SCAP protein. Accordingly, GI3875380 is referred to herein as ceSCAP. Fragments of between 0.2kb to 2kb were produced in regions of interest. Primers

used for each experiment are shown below. Each primer sequence had at either its 3' or 5' end (as indicated below) the T7 RNA polymerase binding site, ATCGATAATACGACTCACTATAGGG (SEQ ID NO:10), which is designated "T7-" below. The remaining nucleotides in each primer sequence are from ceSREBP (SEQ ID NO:1), ceSCAP, or ceS2P, respectively.

SREBP5'A	T7-CCAGCTCAAGGCCCATCAGG
SREBP3'A	T7-TCACTATCCTCATCATCCTC
SREBP5'B	T7-GTACCCGGAACCAATCAATA
SREBP3'B	T7-CTGATGAATTTCATGATAGA

ceSCAP:

SCAP5'A	T7-CAGGACACTCCGCCTAACGA	(SEQ ID NO:11)
SCAP3'A	T7-ACTTACTCGTCAAATTACTC	(SEQ ID NO:12)
SCAP5'B	T7-GTGGCCTCCAGTTGCTCATG	(SEQ ID NO:13)
SCAP3'B	T7-CTTGTATTAGAAAAAAAGTG	(SEQ ID NO:14)
D2013.8S	T7-TGCCGCCCATCCAAAAGCCTGC	(SEQ ID NO:15)
D2013.8A	T7-TATACTTCGGAACCCCAAGTGG	(SEQ ID NO:16)

ceS2P:

S2P5'	T7-GCTCGGTCATGCGTGGGCGG	(SEQ ID NO:17)
S2P3'	T7-TAGCCGCCTCGACAGATTCC	(SEQ ID NO:18)
S2P5'B	T7-CACCGCACGGAAGCCGACGA	(SEQ ID NO:19)
S2P3'B	T7-CTCATTGAGCTGCCCCACAA	(SEQ ID NO:20)

PCR was carried out with 0.5μM each primer and 0.4 μg genomic DNA using the Expand™ PCR Kit (Roche) at the following conditions:

94°C 1 min 15 sec
35 cycles of: 94°C 15 sec
57°C 45 sec
72°C 1 min.

A small fraction of each reaction (2 to 5 μl) was run on a gel to assure that the PCR worked. The rest of each reaction was precipitated and then resuspended in RNase-free water, to serve as the template for production of sense and antisense RNAs. Sense and antisense RNA were transcribed separately from the DNA template using T7 and T3 RNA polymerases (Promega, Madison, WI; RNA production kit, Cat#1300) following the manufacturer's protocol. The resulting RNA samples were ethanol-precipitated and resuspended in 20 μl of RNAse- free TE (10mM tris, 1mM EDTA), followed by 10 μl of RNase free 3X IM annealing buffer (20mM KPO_4 pH7.5, 3mM KCitrate pH 7.5, 2% PEG 6000). The reactions were mixed and incubated at 68°C for 10 minutes and then at 37°C for 30 minutes to anneal the sense and antisense strands. Alternatively, sense and antisense sequences were transcribed together with T7 sites on both strands following the same protocols.

Injection volumes were selected to deliver $0.5x10^6$ to $1x10^6$ molecules of RNA. Injections were delivered to the gonads or the intestinal cavity of *C. elegans*, and were carried out using the methods of Fire *et al.* (Development (1991) 113:503-514).

For germline RNAi, adult animals were microinjected with RNA into either the gonad or intestine using a glass needle mounted on a Medical Systems Corp. (Holliston, MA) PLI-90 injector. For RNAi of larvae, wild type L1 larvae were isolated by first collecting embryos from gravid adults by digestion in 1.25% sodium hypochlorite, 0.25M potassium hydroxide, and then allowing the embryos to hatch overnight in M9 buffer. Equal volumes of larvae in M9 buffer and RNA were mixed in wells of microtiter plates, incubated for 24 hours at 15°C, and then transferred to standard nematode growth plates.

For visualization of lipid, some of the worms were washed off a plate using M9 buffer (per liter: 30gr Na2HPO$_4$, 15g KH$_2$PO$_4$, 2.5g NaCl, 5gr NH$_4$Cl), collected by centrifugation, and resuspended in a 2ng/ml solution of BODIPY™.FL.C12; stock solution is 1mg/ml in ethanol) prepared in M9 buffer. The worms were placed on a benchtop shaker overnight at room temperature to absorb the dye. Images were captured using a fluorescence microscope (Axioplan™, Zeiss, Thornwood, NY) the next day.

Results:

ceSREBP RNAi

Germline ceSREBP (*pin-1*) RNAi produces several visible phenotypes in the progeny of the microinjected animals. The gross phenotype is a fully penetrant larval arrest. Arrested larvae appear to be at the L2 stage based on gonad and cuticle morphology, although their length is more similar to that of L1 stage larvae. Arrested larvae remain motile and feeding for several days at 20°C before dying. Their intestine appears paler, or less darkly pigmented, than wildtype, and this is referred to as the "pale intestine" or "Pin" phenotype.

Morphological defects in ceSREBP RNAi larvae (L1 and L2 stages) are confined to the intestine, where ceSREBP appears to be primarily expressed, and specifically affect three cytoplasmic structures in intestinal cells. First, there is a dramatic reduction in the number and average size of pigmented droplets in the intestine. This reduction of pigmented droplets seems to account for the Pin phenotype observed at low magnification. These droplets likely contain lipid since they stain with dye-labeled fatty acid (BODIPY™-dodecanoic acid) and their number in various developmental stages and mutants correlates with the level of staining with the dye Sudan black in fixed animals.

These observations indicate that ceSREBP is required for formation and/or maintenance of lipid droplets in the intestine, the main lipid storage organ of *C. elegans*.

Second, the gut granules appear larger and more birefringent than in wildtype. Third, many variably sized vesicles appear in the intestine. These vesicles are spherical and transparent; similar vesicles are only rarely observed in wildtype larvae. The vesicles in ceSREBP RNAi larvae are usually each associated with a gut granule, and they show autofluorescence similar in color and intensity to that of gut granules. Since gut granules are thought to be lysosomal structures, the abnormal vesicles in ceSREBP RNAi larvae may also be lysosomal in origin. While many of the intestinal vesicles are immediately visible upon microscopic examination, the number and size of vesicles appears to increase over several minutes of observation, often as larvae begin to show signs of cellular degradation and death. Ultraviolet illumination accelerates this process in ceSREBP RNAi larvae and can also induce formation of similar vesicles in wildtype, although to a lesser extent. These observations may indicate that absorbance of visible or ultraviolet light by the birefringent, autofluorescent gut granules causes damage that induces swelling of lysosomes and synergizes with the effect of ceSREBP RNAi. The larval arrest and morphological defects in the intestine described above are also observed in mutant larvae homozygous for the *pin-1* (ceSREBP) partial deletion allele *ep79* (see Example 4), suggesting that germline RNAi phenocopies the zygotic null phenotype.

ceSREBP RNAi of larvae at the L1 stage results in apparently normal development through the L2 stage, with approximately normal accumulation of intestinal pigmented droplets. However, most larvae arrest at the L3 or L4 stage and fail to maintain their droplets. Arrested larvae, as well as many fully developed adults, show the Pin phenotype and have a thinner body than normal. The number and size of pigmented intestinal droplets is greatly reduced, as observed in earlier stages for germline RNAi. The finding that the Pin phenoptype can be induced by RNAi treatment after terminal differentiation of the intestine indicates that the phenotype is unlikely to be caused by a developmental defect in the intestinal cells. Rather, ceSREBP may be required continuously for proper functioning of the intestine. The pale, thin appearance of ceSREBP RNAi larvae and adults is similar to that of starved animals; however, the RNAi animals display foraging behavior and pump in bacteria through the pharynx into the intestine. These observations suggest that ceSREBP RNAi larvae are defective in digesting and/or metabolizing food. ceSREBP RNAi larvae show greater dispersal away from the food source than wildtype, possibly because they cannot derive nutrients from the bacteria. Transparent intestinal vesicles are observed less frequently with L1 ceSREBP RNAi than with germline ceSREBP RNAi, although most larvae and adults accumulate many vesicles within several minutes of microscope observation under visible or ultraviolet light. Gut granules of the arrested larvae and adults are ofter larger and more birefringent that normal. Adults that display the Pin phenotype

have fewer embryos than normal in their uterus, suggesting reduced fecundity, and some of the embryos show variable developmental defects. Finally, Pin adults often contain large, transparent vacuoles in the anterior half of the intestine. These vacuoles are distinct from the abnormal vesicles observed in larvae, since the vacuoles are irregularly shaped and not autofluorescent, although their origin remains unidentified. ceSREBP RNAi of larvae at the L2 stage results in the same defects as L1 treatment, but mainly in later stages of development. Most animals arrest at the L4 stage or display the adult defects.

The *daf-2 (e1370)* temperature-sensitive mutation (described by Gems *et al.*, Genetics (1998) 150:129-155) produces an opposite phenotype to that of ceSREBP RNAi, a dark intestine (Din) phenotype associated with increased accumulation of pigmented lipid droplets in the intestine. SREBP RNAi can suppress the Din phenotype of *daf-2 (e1370)*, suggesting interaction between the *pin-1*/SREBP pathway and *daf-2*/insulin-like signaling pathway. Specifically, *daf-2 (e1370)* larvae shifted to non-permissive temperature (25°C) at the L1 stage constitutively form dauer larvae with dark intestines. ceSREBP RNAi of these larvae at the L1 stage results in Pin dauers with reduced intestinal lipid droplets. *daf-2 (e1370)* larvae shifted to non-permissive temperature at the L3 stage, after the critical period for commitment to dauer formation, form L4 larvae and adults with dark intestines. If the larvae are also treated with *pin-1* RNAi at the L1 stage, then they can develop a less dark intestine at the L4 and adult stages. *pin-1* does not appear to be strictly epistatic to *daf-2* – rather, double mutants show an intermediate phenotype. Some *pin-1 (ep79)* homozygotes escape larval arrest and can establish semi-viable strains of Pin animals with small, thin bodies and reduced brood size. Double mutants *daf-2 (e1370) pin-1 (ep79)* at 20°C are partially suppressed for all these phenotypes and, in particular, show a less pale intestine. These results suggest that *pin-1* and *daf-2* interact to determine the level of lipid accumulation in the intestine.

ceS2P RNAi

Germline RNAi of the site 2 protease (S2P) homolog results in apparently normal development through the adult stage, however adults show a fully penetrant phenotype exhibiting all the defects observed for *pin-1* larval RNAi (except larval arrest). Specifically, the adult phenotype includes a small, thin body, pale intestine associated with few lipid droplets, abnormally large and birefringent gut granules, large vacuoles in the anterior intestine, fewer embryos in the uterus, and variable developmental defects in some of the embryos. The gut granule defects seem more pronounced than observed for *pin-1* RNAi. The striking similarity of the RNAi phenotypes for ceS2P and *pin-1* strongly suggest that these two genes function in a common genetic pathway. The lack of effect of ceS2P RNAi

on larval development may indicate functional redundancy with an unidentified gene or reduced potency of RNAi for ceS2P compared to *pin-1*.

ceSCAP RNAi

Germline RNAi of the SCAP homologue generates a phenotype similar to ceS2P RNAi in less than 10% of adults. Defective adults display a pale intestine, small and thin body, few embryos in the uterus, and slightly more birefringent gut granules. Germline RNAi of both ceS2P and ceSCAP together produces a fully penetrant phenotype indistinguishable from *pin-1* germline RNAi. This phenotype includes L2-L3 larval arrest, pale intestine associated with few or no intestinal lipid droplets, and abnormally large and birefringent gut granules. These results suggest that both the ceS2P and ceSCAP homologues function in the *pin-1* genetic pathway at all larval and adult stages. If RNAi of ceS2P or ceSCAP produces the null phenotype for these genes, then there must exist other gene activities that can partially substitute for their functions, presumably in proteolytic cleavage at site 2 and 1 analogues, repectively, of PIN-1.

EXAMPLE 4: DOMINANT NEGATIVE ceSREBP PHENOTYPES

A putative dominant negative form of ceSREBP (ceSREBP.DN) was constructed containing amino acids 90-480 of ceSREBP (SEQ ID NO:2). This form lacks the amino-terminal acidic transcriptional activation domain, as well as the C-terminal regulatory part which includes both transmembrane domains. This inactive version which, should not be subject to normal ceSREBP processing, is expected to dimerize with the wild type protein and thereby decrease overall transcriptional activity.

ceSREBP.DN was amplified by PCR using sense primer CeSREBP5'DNSacI (CCCGAGCTCATGCGATTTTCCCCGCCAAACTTTGATC; contained within SEQ ID NO:1), and antisense primer CeSREBP3'CAMfeI (GGGCAATTGCTAAAGGGTAACTTTCGAAGATCCATCTC; contained within SEQ ID NO:1). The fragment was cloned into vector pPD99.52 (Fire Lab) behind the heat shock promoter hsp16/41, which allows temperature-induced activation of the downstream gene.

ceSREBP.DN was injected into N2 worms using standard protocols for *C. elegans* transformation (*Caenorhabditis elegans*: Modern Biological Analysis of an Organism, *supra*) at a concentration of 10 µg/ml plus 100 µg/ml pRF4 rol-6(d) transformation marker, and stable lines displaying the roller phenotype were established.

Misexpression of the ceSREBP.DN transgene was induced by incubating the worms carrying the transgene at 33-34°C. Worms were grown at 20°C before and after the heat-shock. Embryos received a 30 minute heat-shock; larvae and adults received a 2-3 hr

heat-shock. Worms were analyzed for several days after the heat shock under the dissecting microscope, to assess characteristics such as developmental stage, size, pigmentation, mobility (as an indicator of general health), and development of the germ line. Some worms were also analyzed using Nomarski optics on the Zeiss Axioplan™ to assess cellular defects, particularly in the intestine and germ line.

Results:

Results of heat-shock experiments are as follows and are characterized in terms of the phenotypes of the majority, and minority or variable phenotypes:

When embryos are heat-shocked, the majority of animals exhibit slow growth and become small adults with defective germlines and intestinal defects, including reduced lipid content and especially birefringent gut granules. A minority of the animals show embryonic arrest or larval lethality, or become adults with pale intestines.

When L1 larvae are heat-shocked, most animals exhibit slow growth and develop into small adults with defective germlines, a mottled appearance, and intestinal defects. A minority of animals have clear vesicles in their intestines.

When L2 larvae are heat-shocked, the majority of animals develop pale intestines as late larvae. A minority of animals exhibit slow growth, become adults with pale intestines, and/or small adults with defective germlines.

When L3 or L4 larve are heat-shocked, the majority become adults with a mottled appearance and especially birefringent gut granules. A minority of animals exhibit slow growth, become adults with defective germlines, or become very pale and sickly adults.

The majority of heat-shocked adults display no consistent phenotypes, but have various instestinal and germline defects.

The pale intestine phenotype that results from mis-expression of the dominant negative construct is consistent with the pale intestine phenotype that results from ceSREBP RNAi (described in Example 3 above). The construct may be used as a counterscreening reagent, in screens for modifiers of ceSREBP.

EXAMPLE 5: TC1 TRANSPOSON MUTAGENESIS

The goal of this set of experiments was to produce loss-of-function mutations in genes of interest in order to understand the function of their wild-type counterparts.

Library preparation

A Tc1 transposon insertion library comprising 3 sets of 960 cultures was constructed according to published protocols (Zwaal *et al.*, *supra*, and Plasterk, *supra*). Very briefly, 5-10 non- synchronized mut-2(MT3126) animals were cultured on 100 mm peptone plates (2880 plates total) for 12 days. Each culture was then resuspended in M9 medium and

aliquoted into 3 separate tubes, in identical positions, of 3 different racks (each rack holding 96 tubes). Two of the aliquots were frozen for long-term storage, and one lysed for DNA preparation. Lysates were pooled in a 3-dimensional matrix, and their DNA was purified. 10X to 50X dilutions of each DNA prep were used for library screenings by PCR.

Library screening

The library was screened in individual tiers, each library having three tiers, with each tier composed of 1,000 lysates or ~200,000 haploid genomes. Lysates were pooled according to the published protocol. A first dimension screen involved PCR on 8 samples of pooled DNA from ten 96-well plates. A second dimension screen was used to determine which of the ten 96-well plates contained the desired mutant (involves screening of 10 DNA pools). A third dimension screen was used to determine the "address" of a particular mutant (*i.e.*, in which column and row a particular mutant resided - via screening of 12 individual lysates from a single row). First dimension reactions were done in quadruplicate; second and third were done in triplicate.

Two rounds of PCR were performed the first with a pair of gene-specific primers and the second with a pair of Tc1-specific primers. Two different pairs of Tc1 primers were used: one pair pointing outward from the left of the transposon, and the other pair pointing outward from the right (these primer pairs are described in the references cited above).

The first and second round PCR for each dimension was performed in 15 µl total volume using the following in each reaction:

1X PCR buffer provided by the manufacturer (Perkin-Elmer)
1.5 mM $MgCl_2$
0.2 mM dNTPs
0.5 µM each of the Tc1 and the gene-specific primer
0.5 units of Taq Polymerase (Perkin-Elmer)

H_2O to 13 µl for the first round reactions, and to 15 µl for the second round

First and Second dimension: 2 µl of 1:20 diluted DNA was added; 1:10 DNA diluted was added to the third dimension reactions. A small amount of first round reaction was transferred to the second round using a pin replicator. PCR cycling conditions were: 94°C for 3 minutes; then 94 °C for 40 seconds, 58°C for 1 minute, 72°C for 2 minutes for 35 cycles; then 72°C for 2 minutes.

Insertion Screening

The primers used for Tc1 insertion analsysis were as follow:

Tc1 Primers:
Tc1 L1 (round 1, left) CGTGGGTATTCCTTGTTCGAAGCCAGCTAC(SEQ ID NO:21)

Tc1 L2(round 2, left)	TCAAGTCAAATGGATGCTTGAGA	(SEQ ID NO:22)
Tc1 R1 (round 1, right)	TCACAAGCTGATCGACTCGATGCCACGTCG	(SEQ ID NO:23)
Tc1 R2 (round 2, right)	GATTTTGTGAACACTGTGGTGAAGT	(SEQ ID NO:24)

C. elegans SREBP gene-specific primers (each contained within SEQ ID NO:1):

Y47-1 (round 1)	CCCACTCTGTCAAAATCTTCGG
Y47-2 (round 2)	TCAGTGAATAGTGTTGCCGTGC
Y47-4 (round 1)	AGCAATGGAACATATCAACGGG
Y47-3 (round 2)	ACGACCAAGGTTTTCTTTTCCC
Y47-5 (round 1)	TCATTGAGGTATGGTGTGGTGG
Y47-6 (round 2)	GACCTCCACCCATTTTTGTGAG
Y47-8 (round 1)	TGTTGTTTGTGCACAGCATGAG
Y47-7 (round 2)	ACGAGCCCTCAGAACAAAACAG

Results:

Four confirmed Tc1 insertions were found in the ceSREBP gene: one insertion within intron 2 (found using Y47-5/6 and Tc1 R1/R2; address 1D10); one insertion within intron 5 (found using Y47-4/3 and Tc1 L1/L2; address 5D10); one insertion within intron 7 (found using Y47-1/2 and Tc1 R1/R2; address 6D2); and one insertion within intron 8 (found using Y47-1/2 and Tc1 L1/L2; address 1D2). All addresses are from Tier 1 of the Tc1 library described above.

Two of the insertion addresses were chosen for further analysis based on their relatively central location within the SREBP gene: 5D10, located just upstream of the predicted basic helix-loop-helix coding region, and 6D2, located downstream of the two predicted transmembrane domain coding regions.

Identification of insertion animals:

Nematodes from the 6D2 and 5D10 addresses were recovered from frozen stocks representing these addresses; these stocks were made from each culture upon preparation of the library. In order to identify a nematode carrying either a 6D2 or 5D10 insertion, individual surviving nematodes were cloned to individual plates, and after progeny from these nematodes were present on the plates, the parent nematodes were picked into individual wells of a 96-well plate containing 5μl of nematode lysis buffer (100 mM KCl, 20 mM Tris-HCl pH 8.3, 5 mM $MgCl_2$, 0.9% Nonidet P-40, 0.9% Tween-20, 0.02% gelatin, and 400 µg/ml proteinase K). The nematodes were lysed in a PCR machine at 60°C for one hour, followed by 95°C for 15 minutes. 18 µl of a PCR master mix then was added to the crude lysates (to give ~20 µl total reaction volume, assuming evaporation of a portion of the lysate); this mix contained:

1X reaction buffer provided by the manufacturer (Perkin-Elmer)
1.5 mM $MgCl_2$

0.2 mM each dNTP
0.5 µM each gene-specific primer
0.5 units Taq polymerase
to 18 µl per reaction with dH_2O

The PCR reactions were cycled using a program identical to that used for screening the library for the insertions described above. Subsequently, a second round of PCR was performed using the same conditions and primers noted above for the insertion screen, after transferring a small amount of the first round reaction to the second round master mix using a pin replicator. Reactions were run on 1% agarose gels, and gels were analyzed for insertion products identical in size to those observed in the original screen for insertions.

Using this PCR-based screen, a population of nematodes was obtained that is homozygous for the 6D2 insertion. However, since the location of this Tc1 insertion was confirmed to be within an intron, and Tc1 elements are often completely removed along with the intron during splicing of the pre-mRNA, this insertion population was used to identify a deletion in the ceSREBP gene by imprecise excision of the Tc1 element (as described above).

Identification of a Tc1-mediated deletion

In order to obtain a Tc1-mediated deletion in the ceSREBP gene, a small library consisting of 244 cultures of 6D2 insertion nematodes was generated. To create the library, ~5-10 nematodes homozygous for the 6D2 insertion were seeded onto individual plates. After these nematodes had grown, reproduced, and consumed all of the bacteria on these plates, triplicate lysates representing these cultures were created by collecting a sample of nematodes from each plate by washing with a solution of distilled water, and placing the nematodes washed from each plate in one well of a 96-well plate (this was repeated two additional times to create a triplicate set of lysates). Nematodes were lysed by addition of an equal volume of lysis buffer (100 mM KCl, 20 mM Tris-HCl pH 8.3, 5 mM $MgCl_2$, 0.9% Nonidet P-40, 0.9% Tween-20, 0.02% gelatin, and 400 µg/ml proteinase K) followed by incubation at -80°C for 15 minutes, 60°C for 3 hours, and 95°C for 15-30 minutes.

Deletion screening was carried out using a PCR-based approach similar to that used for insertion screening, both of which have been described previously (Zwaal *et al.*, *supra*; and Plasterk, *supra*). Two sets of gene-specific primer pairs were chosen for carrying out a nested PCR strategy such that an outside set was used for the first round of PCR and an inside set was used for the second round of PCR. The second round of PCR was performed to achieve greater specificity in the reaction. The primer sets listed below were chosen since they are ~3.2 kb apart in the ceSREBP genomic sequence (within the typical range for

Tc1 deletion screening), and since they flank either side of the Tc1 insertion in the 6D2 population.

Deletion Screening

ceSREBP gene-specific primers used to identify candidate deletions in ceSREBP were: Y47-1 (round 1), Y47-2 (round 2), Y47-13 (round 1), and Y47-14 (round 2). Primers used in a "specificity test", i.e. a secondary screen for confirming candidate deletions, were: Y47-1 (round 1), Y47-2 (round 2), Y47-4 (round 1), and Y47-3 (round 2). All primers are contained within SEQ ID NO:1:

Y47-1	CCCACTCTGTCAAAATCTTCGG
Y47-2	TCAGTGAATAGTGTTGCCGTGC
Y47-13	GCTTCTTCGGTTACTAGTTAAC
Y47-14	TCAGGAGCATGTTCAGCGACG

The first round PCR reactions were performed using 2 µl of lysate from two of the three sets of lysates, with reactions carried out in a 96-well plate. Each lysate was added to 18 µl of PCR reaction master mix aliquoted into each well:

1X reaction buffer provided by the manufacturer (Perkin-Elmer)
1.5 mM $MgCl_2$
0.2 mM each dNTP
0.5 µM each gene-specific primer
0.5 units Taq polymerase
to 18 µl per reaction with dH_2O

The reactions were carried out in duplicate using the following cycling parameters: 94°C for 3 minutes, then 35 cycles of the following: 94°C for 40 seconds, 55°C for 1 minute, and 72°C for 1 minute. The second round of PCR was performed essentially as above, except that 19.5 µl of the mixture as described for the first round reaction was aliquoted to each reaction. A small amount of first-round reaction products was transferred to the second-round reaction mixtures using a 96-pin replicator. The same temperature cycling sequence was used for the second round as described for the first round.

Products of the second round of PCR were analyzed by electrophoresis in 1% agarose gels. A potential deletion product was observed in both of the reactions, and the putative positive lysate was re-tested by performing duplicate reactions using the relevant lysate from all 3 sets of the library (for a total of six reactions) in two rounds of PCR as described above. The product was gel purified and sequenced directly to confirm the presence of the desired deletion. In addition, in order to confirm that the deletion product obtained was specific for the SREBP region (i.e. not an artifactual result of the PCR), an additional primer set was used in two rounds of PCR as above in a separate set of reactions with all three lysates along with one of the two original primer pairs. This primer set was

chosen such that the PCR product generated would be ~100-300 base-pairs different in size from the original deletion product, resulting in a noticeable shift in size from the original product when analyzed on 1% agarose gels. This part of the screening procedure is termed the "specificity test". Using this procedure to screen the 244 lysates from the 6D2 insertion library with the primers listed above, one deletion of ~2.2 kb within the ceSREBP genomic region was identified, and confirmed by the specificity test (primers used for this test are included in the table above) and by sequence analysis. This deletion begins within intron 6, and ends within exon 9 of the ceSREBP gene. After confirmation, this partial deletion allele was named *pin-1* (*ep79*).

Identification of deletion animals

Following the identification and confirmation of this ~2.2 kb deletion, 192 individual nematodes from the relevant plate were cloned onto separate, fresh plates. When F1 animals were present on the plate, the parent nematodes were placed into buffer present in 96-well plates and lysed as described above. The same primer pairs and cycling conditions used to identify the deletion were used to perform PCR on these animals. Of the192 nematodes screened, one was found by PCR to carry the deletion.

Analysis of mutant phenotypes

Prior to analysis of the SREBP deletion animals, animals carrying the SREBP deletion were outcrossed ten times to a wild-type (N2/Bristol) strain in order to remove extraneous, unrelated mutations induced by the high number of Tc1 elements present in the original mutator strain from which the insertion and deletion in the ceSREBP gene were isolated. Throughout the outcrossing procedure, the SREBP deletion was followed and maintained by analyzing progeny of these crosses by PCR, using the same primers and conditions used for the deletion screen above.

Since reduction or elimination of function mutations often recapitulate phenotypes observed by RNA mediated interference, which in the case of ceSREBP included larval arrest, the deletion mutation was placed in *trans* to a balancer chromosome, and maintained as a heterozygous strain. This is based on the assumption that homozygous deletion mutants would not be able to be propagated themselves if the mutation results in a larval arrest phenotype.

The outcrossed and balanced strain was analyzed for any mutant phenotypes that might arise as a result of the SREBP deletion. It was found that ~25% of the progeny derived from heterozygous SREBP deletion animals (which would correlate to presumptive deletion homozygotes) gave rise to phenotypes observed as the result of SREBP

RNA-mediated interference described in Example 3 above. These phenotypes include: early larval arrest, reduced pigmentation as a result of reduced number of lipid droplets in the intestine, and accumulation of fluid-filled vesicles.

EXAMPLE 6: CLONING OF *DROSOPHILA* S2P

Using BLAST, two EST clones from the Berkeley *Drosophila* Genome Project (BDGP), LD11632 (AA391707) and LD14421(AA439767) were found to have homology with hamster S2P (GI2745731). The sequences were contained in two P1 clones D379 and D380 (AC005465). Primers were used for primer walking to get the full-length DNA sequence. Several more sequencing reactions were performed to produce a complete and unambiguous coverage of the gene which is referred to herein as *Drosophila* S2P (dS2P). The primer sequences below are contained within SEQ ID NO:3.

dS2P SEQUENCING PRIMERS:

(from primer walking)

507:	GGTGAACAAGACAGCTCTTCG
852:	AACGGTGGGAATCACTATGTCAG
1118:	TGATGGTCAGCTACAGTGCTG
186:	TTTCGTGAAGGTGAAATAGCAG

(to resolve ambiguities)

dS2P.s2:	GGTCTTCAGCATAGGATTGG
dS2P.s3:	CACAGTTCGAGTGACATCCC
dS2P.s4:	GTGAGATGGCGCTGCTTTCG
dS2P.s5:	GCACAAGGGTTGTGATGTAG
dS2P.s6:	TACTCAGCCCGGTGTTCTTG

Results:

A full length clone (SEQ ID NO:3) was identified that contained a single open reading frame with an apparent translation start site at nucleotide position 219, and a stop signal at nucleotide position 1745. The predicted polypeptide precursor is 508 amino acids long (SEQ ID NO:4). A search of the PFam and PROSITE databases (Sonnhammer *et al.*, Genomics (1997) 46:200-216; Bairoch *et al.* NAR (1991)19 Suppl:2241-2245; and Hofmann *et al.*, NAR (1999) 27:215-219) revealed seven transmembrane domains and a PDZ domain. The transmembrane domains are located at approximately amino acid residues 4 to 20 (TM1), 82-98 (TM2), 143-159 (TM3), 163-179 (TM4), 208-224 (TM5), 428-444 (TM6) and 478-494 (TM7). The putative PDZ domain is located at approximately amino acid residues 215-285.

The presence of other gene and protein sequences bearing significant homology to *Drosophila* S2P (Fig.2, SEQ ID NO.4) was investigated using the BLAST family of computer programs (Altschul *et al.*, *supra*). The following amino acid sequences were the

most similar: S2P *Homo sapiens* (GI2745733); S2P *Cricetulus griseus* (GI2745731); SP2 metalloprotease, *Homo sapiens* (GI4164134 and GI4164135); putative protein *Arabidopsis thaliana* (GI2982448); conserved protein *Methanobacterium thermoautotrophicum* (GI2622476); and Orf c04034 *Sulfolobus solfataricus* (GI1707806). The most homologous sequence was human S2P (GI2745733) which shared 9 contiguous amino acids at positions 201-207 of SEQ ID NO:4. Amino acid 127 to 501 of SEQ ID NO:4 shares 32 % sequence identity with amino acids 148 to 515 of GI2745733.

EXAMPLE 7: CLONING OF *DROSOPHILA* SCAP

The *Drosophila* SCAP homologue (dSCAP) identified herein, was cloned by PCR based on sequence from a gene prediction and from 5' RACE. BLAST analysis of the hamster SCAP (GI1675220) revealed a genomic P1 clone, DS06954, with regions of high homology. GENSCAN genefinder analysis of this P1 predicted a cDNA that included these homologous regions and was partially covered by ESTs. dSCAP was cloned in overlapping N-terminal and C-terminal fragments with a common HindIII restriction enzyme site.

N-terminal sequence not represented within the gene prediction was obtained by RACE from embryo cDNA prepared with Marathon system (Clontech). A short N-terminal fragment was amplified using non-specific primer AP1 (CCATCCTAATACGACTCACTATAGGGC; SEQ ID NO:25) to the Marathon adaptor and antisense primer dSCAP6 (TCTGGTCCAGCTGCCCGT GTGTTCC; contained within SEQ ID NO:5) contained within the gene prediction and the 5' EST. Amplification conditions were as follows:

1 µl Marathon cDNA
1 µl 10 mM dNTPs
5 µl Klentaq™ buffer
2 µl AP1, 5 µM
2 µl dSCAP6, 5 µM
1 µl Klentaq™ polymerase
38 µl H_2O

94° C 2min	
5 cycles of:	94°C 15 sec
	70°C 4 min
5 cycles of:	94°C 15 sec
	68°C 4 min
25 cycles of:	94°C 15 sec
	62°C 30 sec
	72°C 3 min
72°C 4 min	
12°C hold	

1 µl Amplitaq™ added with an additional incubation for 10 min at 72°C

The major PCR product as determined on a 1% agarose gel was an ~0.7 kb band. This fragment was cloned into the pCRII shuttle vector (Invitrogen) and completely sequenced using M13 forward and reverse primers, and the start codon was identified. Based on the N-terminal sequence identified, a longer N-terminal fragment was amplified from Marathon embryo cDNA using primers dSCAP11(TTGGTATACGGATAGAAATTGG; SEQ ID NO:5) and dSCAP2 (GCGTTTGGGTATTCGTTGCTCC; SEQ ID NO:5). Conditions were as follows:

2 µl Marathon cDNA
5 µl 2 mM dNTPs
5 µl Expand High Fidelity 10X buffer with $MgCl_2$
3 µl dSCAP11, 5 µM
3 µl dSCAP2, 5 µM
0.75 µl Expand High Fidelity enzyme mix
31.25 µl H_2O

94° C 3min
35 cycles of: 94°C 15 sec
50°C 30 sec
72°C 4 min
72°C 4min
12°C hold
1 µl Amplitaq™ added with an additional incubation for 10min at 72°C

The major PCR product as determined on a 1% agarose gel was an ~1.6 kb band. The C-terminal fragment was amplified from embryo 1st strand cDNA using sense primer dSCAP3 (CTCAGTCGCATCCAAAACTGTG; SEQ ID NO:5) and antisense primer dSCAP4 (TTA GGCGCGCCTATTCCTAGGTGCTAGCGAACC; SEQ ID NO:5) made to the predicted cDNA sequence. Amplifications conditions were as follows:

2 µl 1st strand cDNA
5 µl 2 mM dNTPs
5 µl Expand High Fidelity 10X buffer with $MgCl_2$
3 µl dSCAP3, 5 µM
3 µl dSCAP4, 5 µM
0.75 µl Expand High Fidelity enzyme mix
31.25 µl H_2O

94°C 3min
15 cycles of: 94°C 15 sec
60°C 30 sec
72° C 2 min
20 cycles of: 94°C 15 sec
60°C 30 sec
72°C 2min + 20sec/cycle
72°C 5 min
4° C hold
1 µl Amplitaq™ added with an additional incubation for 10min at 72°C

The major PCR product as determined on a 1% agarose gel was an ~2.2 kb band.

Both N-terminal and C-terminal fragments were cloned into pCRII and completely sequenced in both directions.

Results

A full-length clone was identified that contained a single open reading frame with an apparent translational initiation site at nucleotide position 73 and a stop signal at nucleotide position 3786 (SEQ ID NO:5). The predicted polypeptide precursor is 1237 amino acids long (SEQ ID NO:6). Additional features include:

1) A Ribosomal RNA adenine dimethylase at nucleotides 667 to 703 (amino acid residues 198 to 210);

2) Four G-beta (GB) repeat WD domains: GB1 at nucleotides 2509 to 2617, corresponding to amino acid residues 812 to 848; GB2 at nucleotides 3080 to 3196, corresponding to amino acid residues 1005 to 1041; GB3 at nucleotides 3208 to 3325, corresponding to amino acid residues 1045 to 1084; and GB4 at nucleotides 3337 to 3445, corresponding to amino acid residues 1088 to 1124;

3) Six predicted transmembrane (TM) domains. TM1 at nucleotides 991 to 1039, corresponding to amino acid residues 306 to 322; TM2 at nucleotides 1117 to 1165, corresponding to amino acid residues 348 to 364; TM3 at nucleotides 1180 to 1228, corresponding to amino acid residues 369-385; TM4 at nucleotides 1366 to 1414, corresponding to amino acid residues 431-447; TM5 at nucleotides 1753 to 1801, corresponding to amino acid residues 560 to 576; and TM6 at nucleotides 2353 to 2401, corresponding to amino acid residues 760 to 776.

The presence of other gene and protein sequences bearing significant homology to dSCAP (SEQ ID NO:5) was investigated using BLAST (Altschul *et al.*, *supra*) against nucleotide databases. This revealed that dSCAP is covered by two genomic clones from BDGP: DS06954 (P1 D338), and DS05325 (P1 D340). The accession number for the two clones is AC007121. Other sequences bearing nucleotide homology with dSCAP are human mRNA for KIAA0199 gene (GI 1228046), and *Cricetulus griseus* SCAP mRNA (GI 1228046). At the protein level, dSCAP shares homology with the following sequences: *C. elegans* predicted SCAP D2013.8 (GI 642180), *Homo sapiens* KIAA0199 gene (GI 1228047), *Cricetulus griseus* SCAP (GI 1675220), and is similar to the transmembrane domain of HMGCOA (GI 3875380).

EXAMPLE 8: TRANSGENIC *DROSOPHILA* MISEXPRESSING SREBP

The wild-type *Drosophila* SREBP (dSREBP) (HLH106) gene was cloned by PCR. The coding sequence of the gene was amplified in overlapping N-terminal and C-terminal

regions from a *Drosophila* adult cDNA library (Stratagene, cat #936603). Primers used to amplify the N-terminal region were sense primer HLH106.1 (AATGGACACGACACTGATGAAC; SEQ ID NO:7) and antisense primer HLH106.2 (AGCCATGTTGCTTGCGAATAGT; SEQ ID NO:7). Primers used to amplify the C-terminal regions were sense primer HLH106.3 (AAACAGGCGCTGGCATCTGCAC; SEQ ID NO:7) and antisense primer HLH106.4 (GGCGCGCCCACGTTCGTGCCACTTATTATGTA; SEQ ID NO:7). The fragments were spliced together using the common restriction site SacII.

In addition to the wild-type gene, one putative constitutively active form dSREBP (dSREBP.CA) and three putative dominant negative forms were engineered for misexpression in *Drosophila*. All were designed based on precedents in mammalian SREBP research (reviewed by Brown and Goldstein, *supra*). These constructs, as well as the wild-type gene may be used both as screening or counterscreening reagents, and as devices to further elucidate the function of SREBP in *Drosophila*.

Sequences of all fragments were verified. All constructs were cloned into pExPress-UAS . pExPress is a vector designed specifically for misexpression of genes in transgenic *Drosophila*. This vector was derived from pGMR (Hay *et al.*, Development (1994) 120:2121-2129). The vector is 9Kb long, and contains: an origin of replication for *E. coli*; an ampicillin resistance gene; P element transposon 3' and 5' ends to mobilize the inserted sequences; a White marker gene; an expression unit comprising the TATA region of hsp70 enhancer and the 3'untranslated region of α-tubulin gene. The expression unit contains a first multiple cloning site (MCS) designed for insertion of an enhancer and a second MCS located 500 bases downstream, designed for the insertion of a gene of interest. DNA constructs are cloned into the EcoR1 and/or EcoR1/AscI sites of the second MCS.

Fragments cloned into pExPress-UAS were injected into yw *Drosophila* embryos using standard protocols for *Drosophila* transformation (Rubin and Spradling, *supra*). A variety of GAL4 driver lines were used to drive mis-expression of the transgenes. Driver lines glass, sevenless, Kruppel, Rhomboid, 2677, and 1878 are available from the University of Indiana (http://flybase.bio.indiana.edu). Lines T93, T113, and T155 were kindly provided by Tian Xu (Yale University School of Medicine, New Haven, CT, USA) Descriptions of the larval expression patterns of the GAL 4 are presented in Table III.

TABLE III

GAL4-Driver	Larval Expression Pattern
EYE	
3X glass (GMR)	Photoreceptor cells, very strong expression
2X sevenless	R7 photoreceptor cells
2677	Transiently, during eye development
FAT BODY	
T93	Fat body, wing and eye discs, brain, salivary glands
T113	Fat body, wing and eye discs, salivary glands
T155	Fat body, wing and eye discs, brain, salivary glands
GUT/GENERAL	
1878	Ubiquitous (fat body, gut, discs, trachea, brain, etc.)
Kruppel	General gut, fat body, brain and segmental neurons, salivary glands
Rhomboid	Whole gut, segmental nerves, salivary glands, minor fat body and salivary gland staining

The putative activated form, dSREBP.CA, contains amino acids 1-448 of dSREBP (SEQ ID NO:8) and lacks the C-terminal regulatory region, including the membrane-spanning domains, and thus should require no processing to activate transcriptional targets. dSREBP.CA was amplified by PCR from a clone of wild-type dSREBP using sense primer HLH106.1 (AATGGACACGACACTGATGAAC; SEQ ID NO:7) and antisense primer HLH106.CA (CTAGCGAGAGTGGGTGGCCATGC; SEQ ID NO:7). The observed phenotypes for this construct under various driver lines are presented in Table IV. The phenotypes exhibited by expression in the fat body is evidence that the dSREBP transgene exerts metabolic effects.

TABLE IV

GAL4-Driver	Phenotype
Line #1- No driver	Bristles (macrochaete) shortened, often missing
Line #2 - No driver	No phenotype
EYE	
3X glass (GMR)	**Line 1:** Lethality, embryonic or larval **Line 2:** Strong rough eye
2X sevenless	**Lines 1 & 2:** Lethality, early pupal
2677	**Line 1:** Lethality, embryonic or larval **Line 2:** Rough and reduced eye
FAT BODY	
T93	**Lines 1 & 2:** Reduced male viability, reduced female fertility, adults with caved-in abdomens and starved appearance, persistence of the larval fat body in adults, short life spans (all w/variable penetrance)
T113	**Line 1:** Mostly pupal lethal, most survivors are female and have the abdomen phenotype of T93.
T155	**Lines 1 & 2:** Larval lethal; a few escapers appear normal.
GUT/GENERAL	
1878	**Lethal- embryo or larvae**
Kruppel	**Lethal- embryo or larvae**
Rhomboid	**Lethal- embryo or larvae**

Two putative dominant negative forms of dSREBP, "dSREBP.DNr (*Dominant Negative regulated*)" and "dSREBP.DNur (*Dominant Negative unregulated*)" lack the amino-terminal acidic domain and should be transcriptionally inactive. They are expected to act by competing with the wild-type protein in dimerization to make transcriptionally inactive dimers. They differ by the inclusion of the C-terminal regulatory region. dSREBP.DNr includes the full regulatory region and should be active only in conditions in which dSREBP is cleaved from the ER membrane; it contains amino acids 75-1113 of dSREBP. dSREBP.DNr lacks the regulatory region and should not require processing; it may therefore be a more potent inhibitor of transcription. dSREBP.DNur contains amino acids 75-448 of dSREBP.

The 5' part of dSREBP.DNr was amplified by PCR from a clone containing the 5' of the dSREBP, using sense primer HLH106.DN (CGCAATGTCCGTCGAGCAACAGCCGCAC; SEQ ID NO:7) and antisense primer HLH106.2 (AGCCATGTTGCTTGCGAATAGT; SEQ ID NO:7). This fragment was spliced together with an overlapping clone containing the 3' part

of dSREBP using the common restriction site SacII. dSREBPur was similarly amplified using sense primer HLH106.DN and antisense primer HLH106.CA.

The third putative dominant negative "p450/dSREBP" is expected to act through interaction with SCAP. p450/dSREBP contains the C-terminal regulatory region of dSREBP (amino acids 521-1113), fused to the mammalian cytochrome p450 transmembrane domain, which acts to anchor the protein in the ER. The p450 transmembrane, with a 5', in-frame ATG, was generated by annealing two complementary oligonucleotides. Sense-strand oligonucleotide was p450.1 (CTGGAATTCAACATGGATCCAGTGGTGGTGCTGGGACTCTGCCTCTCCTGCTTGCTTCTCCTTT CACTCTGGAAGCAGAGCTATGGAGGAGGAAAGCTT; SEQ ID NO:26). Antisense-strand oligo was p450.2 (AAGCTTTCCTCCTCCATAGCTCTGCTTCCAGAGTGAAAGGAGAAGCAAGCAGGAGAGGCAGAGT CCCAGCACCACCACTGGATCCATGTTGAATTCCAGAGCT; SEQ ID NO:27). This transgene should act by titrating out dSCAP, leaving less available for the processing of wild-type dSREBP. Table V summarizes the observed phenotypes for the dominant negative constructs under various driver lines.

TABLE V

Line	GAL4-Driver	Phenotype
	EYE	
Wild Type	3X glass (GMR)	Rough eye; disorganization of ommatidia
	2X sevenless	Some very mild roughness in posterior eye
	2677	Some very mild roughness in posterior eye
	EYE	
dSREBP.DNr	3X glass (GMR)	Mild rough eye; disorganization of ommatidia
	2X sevenless	No phenotype
	2677	Some very mild roughness in posterior eye
dSREBP.DNur	3X glass (GMR)	Strong rough eye
	2X sevenless	No phenotype
	2677	Some very mild roughness in posterior eye
	GUT/GENERAL	
	Kruppel	Lethality, embryonic or larval
p450/dSREBP	**EYE**	
	3X glass (GMR)	Strong rough eye, fused ommatidia

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Such modifications are intended to fall within the scope of the appended claims. The disclosure of each reference cited herein, including patents and other references, is hereby incorporated herein by reference in its entirety.

WHAT IS CLAIMED IS:

1. An animal that is a fly or nematode that has been genetically modified to express or mis-express an SREBP pathway protein, or the progeny of said animal that has inherited said SREBP pathway protein expression or mis-expression.

2. The animal of Claim 1 that has been genetically modified by a method selected from the group consisting of transposon insertion mutagenesis, double-stranded RNA interference, and chemical mutagenesis.

3. The animal of Claim 1 wherein a heterologous promoter drives expression or mis-expression of said SREBP pathway protein.

4. The animal of Claim 3 wherein said promoter is selected from the group consisting of tissue-specific promoters, developmental-specific promoters, and inducible promoters.

5. The animal of Claim 4 wherein said animal is a fly and said promoter is selected from the group consisting of *sevenless*, *eyeless*, *glass*, *dpp*, heat shock, tTA-responsive, GAL4-responsive, and *vestigal*.

6. The animal of Claim 1 wherein said SREBP pathway protein is encoded by an SREBP pathway nucleic acid sequence linked to a nucleic acid sequence that encodes one or more selectable markers that allows detection of expression of said SREBP pathway protein.

7. The animal of Claim 1 wherein said expression or mis-expression of said SREBP pathway protein results in an identifiable phenotype.

8. The animal of Claim 1 wherein said SREBP pathway protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, 6, and 8, or a functionally-active fragment thereof.

9. The animal of Claim 8 wherein said SREBP pathway protein is encoded by part or all of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, and 7.

10. The animal of Claim 7 wherein said nematode is heterozygous for deletion of SREBP.

11. The animal of Claim 7 wherein said animal is a nematode and said identifiable phenotype is a pale intestine phenotype or other intestinal defect.

12. A method for studying lipid metabolism comprising detecting the phenotype caused by the expression or mis-expression of said SREBP pathway protein in the animal of Claim 1.

13. The method of Claim 12 additionally comprising observing a second animal having the same genetic modification as the animal of Claim 1 which causes said expression or mis-expression of said SREBP pathway protein, and wherein said second animal additionally comprises a mutation in a gene of interest, wherein differences, if any, between the phenotype of the animal of Claim 1 and the phenotype of the second animal identifies the gene of interest as capable of modifying the function of the gene encoding said SREBP pathway protein.

14. The method of Claim 13 wherein said gene of interest is implicated in cholesterol or fatty acid biosynthesis or metabolism.

15. The method of Claim 13 wherein said animal is a nematode and wherein said phenotype is a pale intestine phenotype or other intestinal defect indicative of abnormalities in lipid biosynthesis or metabolism.

16. The method of Claim 13 wherein said animal is a nematode and wherein said method includes staining said nematode *in vivo* with a fluorescently-labelled fatty acid conjugate to measure lipid content within said nematode.

17. The method of Claim 16 wherein said fluorescently-labelled fatty acid conjugate is a BODIPY™-fatty acid conjugate.

18. The method of Claim 13 additionally comprising administering one or more compounds to said animal or its progeny and observing any changes in lipid content of said animal or its progeny.

19. A method for determining the lipid content of a living nematode comprising contacting said nematode with a BODIPY™ fatty acid conjugate to stain lipid and measuring fluorescence as an indication of lipid content.

20. The method of Claim 19 which is used in combination with a genetic screen for detection of mutations that affect lipid content.

21. The method of Claim 19 that additionally includes administering one or more compounds to said nematode or its progeny and observing any effect said compound has on lipid content.

22. An isolated nucleic acid molecule of less than 15 kb comprising a nucleic acid sequence selected from the group consisting of:
A) a nucleic acid sequence that encodes a polypeptide comprising at least 10 contiguous amino acids of the sequence of any one of SEQ ID NO:2, 4 and 6; and
B) a nucleic acid sequence that encodes a polypeptide comprising at least 8 contiguous amino acids of residues 335 to 428 of SEQ ID NO:2.

23. The isolated nucleic acid molecule of Claim 22 that hybridizes under appropriate conditions to a nucleic acid sequence selected from the group consisting of SEQ ID NOs:1, 3 and 5.

24. The isolated nucleic acid molecule of Claim 23 wherein said appropriate conditions comprise hybridization at 34°C in a buffer comprising 6X SSC / 0% formamide and a wash at 45°C in a buffer comprising 2X SSC.

25. A vector comprising the nucleic acid molecule of Claim 22.

26. A host cell comprising the vector of Claim 25.

27. The host cell of Claim 26 wherein said cell is a yeast cell.

28. A process for producing an SREBP pathway protein comprising culturing the host cell of Claim 26 under conditions suitable for expression of said SREBP pathway protein and recovering said protein.

29. An isolated SREBP pathway protein produced by the process of Claim 28.

30. The isolated SREBP pathway protein of Claim 29 which is joined at its amino- or carboxy-terminus via a peptide bond to an amino acid sequence of a different protein.

31. A nucleic acid molecule comprising a *C. elegans* SREBP promoter operably-linked to a heterologous gene, wherein said SREBP promoter is derived from nucleotides 66,719-71,265 of Y47D3 (GI:3646936).

32. The nucleic acid of Claim 31, wherein said heterologous gene encodes a selectable or detectable marker.

33. A method of detecting a candidate molecule that binds to a polypeptide comprising SEQ ID 2, 4, or 6 comprising:

(a) contacting said polypeptide with one or more candidate molecules under conditions conducive to binding; and

(b) detecting any binding that occurs between the candidate molecules and said polypeptide.

ABSTRACT OF THE INVENTION

Drosophila melanogaster and *C. elegans* that have been genetically modified to express or mis-express proteins involved in the sterol regulatory element binding protein (SREBP) pathway are described. These genetically modified animal models have identifiable phenotypes that make them useful in assays for studying lipid metabolism, other genes implicated in lipid metabolism, and compounds capable of modulating lipid metabolism pathways. Methods for studying lipid metabolism in living nematodes using fluorescently-labelled fatty acid conjugates, such BODIPY™ fatty acid conjugates, are also described. Novel SREBP pathway nucleic acid and protein sequences are also described.

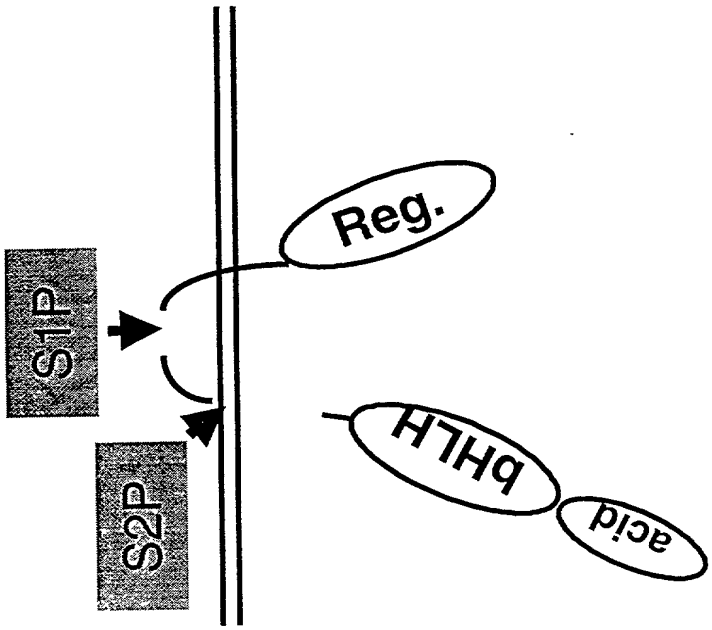


FIG. 1B

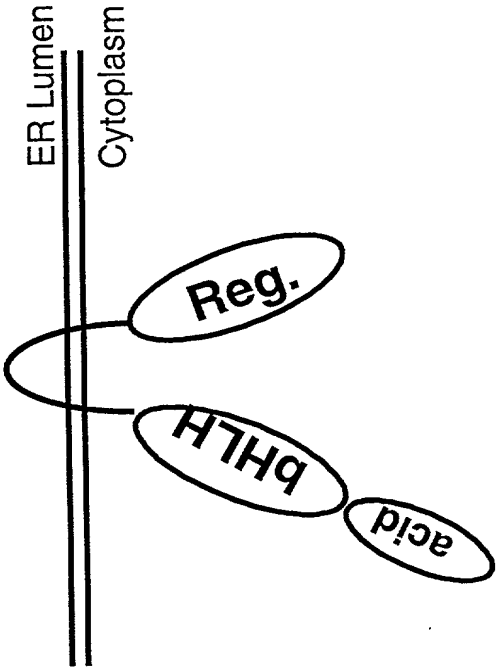


FIG. 1A

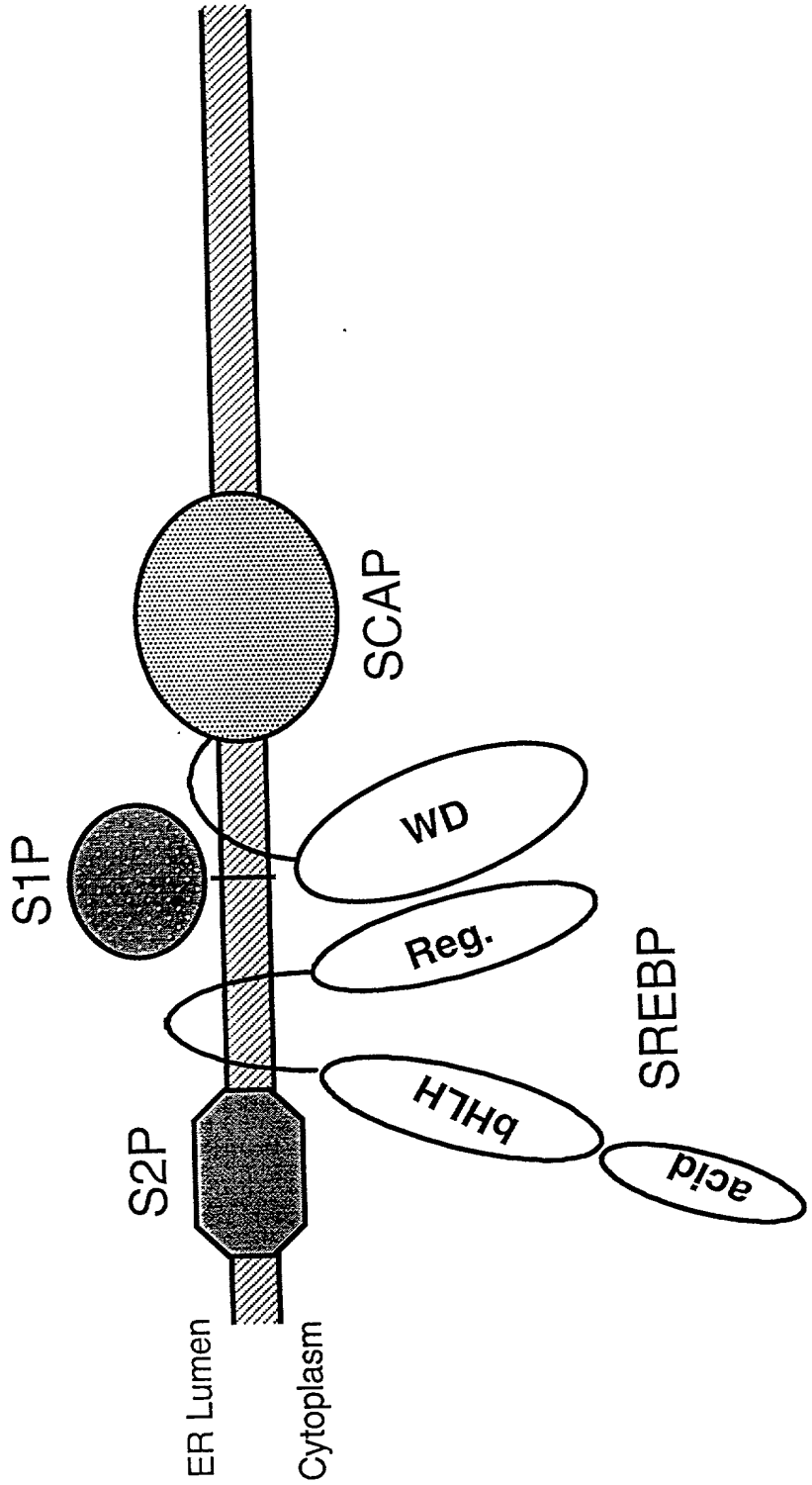


FIG. 2

GGTTTAATTACCCAAGTTTGAGAATGAACGAAGAATTCGAGGGAGACGTC 50
CCAAATTAATGGGTTCAAACTCTTACTTGCTTCTTAAGCTCCCTCTGCAG

CCTATGTCGGATCCGTTTCTCTCATTGGTCACAAAATTGGATGATATTGC 100
GGATACAGCCTAGGCAAAGAGAGTAACCAGTGTTTTAACCTACTATAACG

GCCATTTCCAAATAACGACCCGCTCGATTTTGACATGGAGCACAACTGGC 150
CGGTAAAGGTTTATTGCTGGGCGAGCTAAAACTGTACCTCGTGTTGACCG

AAGAGCCCGGACCATCACAACAACCGGATCCATCAATTCCCGGAAATCAA 200
TTCTCGGGCCTGGTAGTGTTGTTGGCCTAGGTAGTTAAGGGCCTTTAGTT

CACAGTCCGCCACAGGAATATTATGATATTGATGGTCAACGAGACGTAAG 250
GTGTCAGGCGGTGTCCTTATAATACTATAACTACCAGTTGCTCTGCATTC

CACCTTACACTCCCTGCTCAACCACAACAACGACGACTTCTTCTCAATGC 300
GTGGAATGTGAGGGACGAGTTGGTGTTGTTGCTGCTGAAGAAGAGTTACG

GATTTTCCCCGCCAAACTTTGATCTCGGCGGAGGCCGTGGACCTTCTCTA 350
CTAAAAGGGGCGGTTTGAAACTAGAGCCGCCTCCGGCACCTGGAAGAGAT

GCCGCCACCCAACAATTATCTGGAGAAGGTCCTGCAAGTATGCTTAACCC 400
CGGCGGTGGGTTGTTAATAGACCTCTTCCAGGACGTTCATACGAATTGGG

CTTACAAACATCTCCACCAAGTGGAGGTTACCCCCCGGCAGATGCCTACA 450
GAATGTTTGTAGAGGTGGTTCACCTCCAATGGGGGGCCGTCTACGGATGT

GACCTCTATCACTTGCTCAACAACTCGCCGCGCCAGCGATGACTCCACAT 500
CTGGAGATAGTGAACGAGTTGTTGAGCGGCGCGGTCGCTACTGAGGTGTA

CAGGCAGCGTCGCTTTTTGTTAATACTAATGGAATTGATCAAAAGAATTT 550
GTCCGTCGCAGCGAAAAACAATTATGATTACCTTAACTAGTTTTCTTAAA

CACTCATGCAATGCTATCTTCACCACACCATACCTCAATGACTTCTCAAC 600
GTGAGTACGTTACGATAGAAGTGGTGTGGTATGGAGTTACTGAAGAGTTG

CATATACAGAAGCCATGGGACATATCAACGGGTACATGTCTCCATACGAC 650
GTATATGTCTTCGGTACCCTGTATAGTTGCCCATGTACAGAGGTATGCTG

CAAGCTCAAGGCCCATCAGGACCATCATATTACTCACAACACCATCAATC 700
GTTCGAGTTCCGGGTAGTCCTGGTAGTATAATGAGTGTTGTGGTAGTTAG

TCCACCACCTCATCACCACCATCACCACCCGATGCCAAAAATCCATGAGA 750
AGGTGGTGGAGTAGTGGTGGTAGTGGTGGGCTACGGTTTTTAGGTACTCT

ACCCTGAACAAGTGGCATCTCCATCGATTGAAGATGCTCCAGAGACGAAA 800
TGGGACTTGTTCACCGTAGAGGTAGCTAACTTCTACGAGGTCTCTGCTTT

Express Mail Label No. EV339904925US

FIG. 3A

```
CCAACTCATTTGGTTGAACCACAAAGTCCAAAAAGCCCGCAGAATATGAA      850
GGTTGAGTAAACCAACTTGGTGTTTCAGGTTTTTCGGGCGTCTTATACTT

AGAGGAGCTTCTTCGGTTACTAGTTAACATGTCTCCGAGTGAAGTTGAAC      900
TCTCCTCGAAGAAGCCAATGATCAATTGTACAGAGGCTCACTTCAACTTG

GGTTAAAGAATAAAAAATCAGGAGCATGTTCAGCGACGAATGGGCCATCG      950
CCAATTTCTTATTTTTTAGTCCTCGTACAAGTCGCTGCTTACCCGGTAGC

AGGAGTAAGGAGAAGGCGGCGAAGATTGTGATTCAGGAGACAGCGGAAGG     1000
TCCTCATTCCTCTTCCGCCGCTTCTAACACTAAGTCCTCTGTCGCCTTCC

GGATGAAGATGAGGATGATGAGGATAGTGATTCCGGGGAGACTATGTCTC     1050
CCTACTTCTACTCCTACTACTCCTATCACTAAGGCCCCTCTGATACAGAG

AGGGAACTACTATTATTGTTCGAAGACCAAAAACCGAGCGTCGTACGGCA     1100
TCCCTTGATGATAATAACAAGCTTCTGGTTTTTGGCTCGCAGCATGCCGT

CACAATCTCATCGAAAAGAAGTATAGATGCTCAATAAATGATCGAATTCA     1150
GTGTTAGAGTAGCTTTTCTTCATATCTACGAGTTATTTACTAGCTTAAGT

ACAGCTGAAAGTACTTTTGTGTGGGGATGAAGCTAAGCTTTCAAAATCGG     1200
TGTCGACTTTCATGAAAACACACCCCTACTTCGATTCGAAAGTTTTAGCC

CAACACTACGACGGGCTATTGAACATATCGAGGAGGTTGAACACGAGAAT     1250
GTTGTGATGCTGCCCGATAACTTGTATAGCTCCTCCAACTTGTGCTCTTA

CAGGTGTTGAAGCATCATGTTGAACAAATGAGAAAGACACTGCAGAATAA     1300
GTCCACAACTTCGTAGTACAACTTGTTTACTCTTTCTGTGACGTCTTATT

TCGATTACCGTACCCGGAACCAATTCAATACACTGAATACTCTGCCCGAT     1350
AGCTAATGGCATGGGCCTTGGTTAAGTTATGTGACTTATGAGACGGGCTA

CACCCGTCGAATCATCTCCTTCTCCACCTAGAAATGAGAGAAAACGATCA     1400
GTGGGCAGCTTAGTAGAGGAAGAGGTGGATCTTTACTCTCTTTTGCTAGT

CGAATGAGCACAACGACTCCTATGAAGAATGGAACTAGAGATGGATCTTC     1450
GCTTACTCGTGTTGCTGAGGATACTTCTTACCTTGATCTCTACCTAGAAG

GAAAGTTACCCTTTTTGCGATGCTCCTAGCAGTTCTGATTTTTAATCCGA     1500
CTTTCAATGGGAAAAACGCTACGAGGATCGTCAAGACTAAAAATTAGGCT

TTGGATTGCTCGCTGGAAGTGCGATATTCTCAAAAGCCGCTGCAGAAGCT     1550
AACCTAACGAGCGACCTTCACGCTATAAGAGTTTTCGGCGACGTCTTCGA

CCGATTGCCTCCCCGTTCGAGCATGGAAGAGTGATTGATGACCCGGATGG     1600
GGCTAACGGAGGGGCAAGCTCGTACCTTCTCACTAACTACTGGGCCTACC
```

FIG. 3B

```
AACTAGCACTCGGACGCTTTTCTGGGAAGGGAGTATCATCAATATGAGCT    1650
TTGATCGTGAGCCTGCGAAAAGACCCTTCCCTCATAGTAGTTATACTCGA

ATGTCTGGGTGTTCAACATCTTAATGATCATATATGTGGTTGTCAAACTG    1700
TACAGACCCACAAGTTGTAGAATTACTAGTATATACACCAACAGTTTGAC

CTGATCCATGGTGACCCTGTTCAAGACTTCATGTCCGTTTCATGGCAGAC    1750
GACTAGGTACCACTGGGACAAGTTCTGAAGTACAGGCAAAGTACCGTCTG

TTTTGTGACGACTCGAGAGAAGGCGAGAGCCGAGTTGAACTCTGGAAATT    1800
AAAACACTGCTGAGCTCTCTTCCGCTCTCGGCTCAACTTGAGACCTTTAA

TGAAAGATGCTCAGAGAAAGTTCTGCGAGTGTCTTGCAACGTTGGATCGA    1850
ACTTTCTACGAGTCTCTTTCAAGACGCTCACAGAACGTTGCAACCTAGCT

TCGCTTCCATCACCGGGGGTTGATTCGGTGTTTTCGGTTGGCTGGGAATG    1900
AGCGAAGGTAGTGGCCCCCAACTAAGCCACAAAAGCCAACCGACCCTTAC

CGTTCGACATCTTTTGAATTGGTTGTGGATCGGGAGATACATCGCAAGAA    1950
GCAAGCTGTAGAAAACTTAACCAACACCTAGCCCTCTATGTAGCGTTCTT

GGCGCAGGTCCACCACGAAGCCTGTCTCAGTCGTTTGTAGGAGTCATGCG    2000
CCGCGTCCAGGTGGTGCTTCGGACAGAGTCAGCAAACATCCTCAGTACGC

CAGACTGCAGTTCTCTATCATGAAATTCATCAGCTCCATCTAATGGGTAT    2050
GTCTGACGTCAAGAGATAGTACTTTAAGTAGTCGAGGTAGATTACCCATA

CACTGGAAACTTCGAAGACACCTATGAACCATCCGCCCTAACGGGCCTCT    2100
GTGACCTTTGAAGCTTCTGTGGATACTTGGTAGGCGGGATTGCCCGGAGA

TCATGTCCCTCTGTGCAGTAAACCTTGCTGAAGCTGCCGGAGCATCAAAC    2150
AGTACAGGGAGACACGTCATTTGGAACGACTTCGACGGCCTCGTAGTTTG

GACGGACTTCCACGCGCCGTCATGGCTCAGATCTACATTTCTGCATCCAT    2200
CTGCCTGAAGGTGCGCGGCAGTACCGAGTCTAGATGTAAAGACGTAGGTA

CCAATGCCGTTTGGCTCTTCCGAACCTACTCGCACCATTCTTCTCGGGAT    2250
GGTTACGGCAAACCGAGAAGGCTTGGATGAGCGTGGTAAGAAGAGCCCTA

ACTTTTTACGAAGAGCTCGAAGGCACGTGCGTCGAGCTCCGGAGCACTCG    2300
TGAAAAATGCTTCTCGAGCTTCCGTGCACGCAGCTCGAGGCCTCGTGAGC

GTGTCCCATTTGTTATGGATCTTCCATCCAGCGACAAGAAAGTTCATGTC    2350
CACAGGGTAAACAATACCTAGAAGGTAGGTCGCTGTTCTTTCAAGTACAG

AGATGCGAAAAGGTTGGAGCATGTGTTGAGCTCGAAGCAGAAGCAGTTGA    2400
TCTACGCTTTTCCAACCTCGTACACAACTCGAGCTTCGTCTTCGTCAACT
```

FIG. 3C

```
GATTTGGGTCTTTTGTGGAAGATGAGCAATTATCCCCACTTGCTCGAATC      2450
CTAAACCCAGAAAACACCTTCTACTCGTTAATAGGGGTGAACGAGCTTAG

CGAACAACGCTGAAAGTGTACCTACTCTCCAAACTTGTACAGGAACTTGT      2500
GCTTGTTGCGACTTTCACATGGATGAGAGGTTTGAACATGTCCTTGAACA

CGGTGGTGACGAGATCTTTACAAAAAATGTGGAACGCATCCTAAATGACA      2550
GCCACCACTGCTCTAGAAATGTTTTTTACACCTTGCGTAGGATTTACTGT

ATGACCGTCTCGATGATGAAGTAGACGTGGTTGATGTTTCAAGACTTTTG      2600
TACTGGCAGAGCTACTACTTCATCTGCACCAACTACAAAGTTCTGAAAAC

GTGACAATTTCAACGCAGTGCGCTGCCATTTTGACTAATGAGAAGGATGA      2650
CACTGTTAAAGTTGCGTCACGCGACGGTAAAACTGATTACTCTTCCTACT

GTCAGCGAAATTCGGAACCTGGATCTCTCGAAACGGAGATGCTTGTTGCA      2700
CAGTCGCTTTAAGCCTTGGACCTAGAGAGCTTTGCCTCTACGAACAACGT

CATGGTGGACGCACGTTCTGACATGTGGAATCTATTGGAGGAGTAACAAG      2750
GTACCACCTGCGTGCAAGACTGTACACCTTAGATAACCTCCTCATTGTTC

AATGAGCTGGCACGGCAACACTATTCACTGATCAGGAACTGTCCGCCGAA      2800
TTACTCGACCGTGCCGTTGTGATAAGTGACTAGTCCTTGACAGGCGGCTT

GATTTTGACAGACAATCTGGGTTTGGCGGTTGGCCACGCGTTGTGTGCTC      2850
CTAAAACTGTCTGTTAGACCCAAACCGCCAACCGGTGCGCAACACACGAG

GCAAGATTTGCATAGATGACCGAGATTCCCCGAAAGTCAGTCAATACGTG      2900
CGTTCTAAACGTATCTACTGGCTCTAAGGGGCTTTCAGTCAGTTATGCAC

TGCATTCACACAAAGAAGTCGCTCGAATCCCTCCGACTATTCTCCACATC      2950
ACGTAAGTGTGTTTCTTCAGCGAGCTTAGGGAGGCTGATAAGAGGTGTAG

ATCGCGAGCATCAGGTGTGGTGTCTGGAATTCAGGAAGGTACACGCCGAA      3000
TAGCGCTCGTAGTCCACACCACAGACCTTAAGTCCTTCCATGTGCGGCTT

TGGCCTACGAATGGATTATGAACTCGCTGCTCGACGCGTGGCGTTCCAAT      3050
ACCGGATGCTTACCTAATACTTGAGCGACGAGCTGCGCACCGCAAGGTTA

CTATTCGCATCGAAACCCTACTGGACACAAAGCTTCAAGGGACAATCCAC      3100
GATAAGCGTAGCTTTGGGATGACCTGTGTTTCGAAGTTCCCTGTTAGGTG

GTTTAGTACGCTTTATCAAGAGGCGTATAATCATTATGCGATTATTAATG      3150
CAAATCATGCGAAATAGTTCTCCGCATATTAGTAATACGCTAATAATTAC

GGACAAGGGGAGATTGTTGGAGACTATTTGTCTACGAGCTCACGTGCCGA      3200
CCTGTTCCCCTCTAACAACCTCTGATAAACAGATGCTCGAGTGCACGGCT
```

FIG. 3D

```
ATGCTCAACGGAGCCAACCCACAAGCCACGTGGTCAGGCGYCCGACGCGT     3250
TACGAGTTGCCTCGGTTGGGTGTTCGGTGCACCAGTCCGCRGGCTGCGCA

TCGATCTACAAAAATGGACGCGGTCCGAGGAAGAGTGAGCATGCGACGCT     3300
AGCTAGATGTTTTTACCTGCGCCAGGCTCCTTCTCACTCGTACGCTGCGA

CGGCTCAACCGGACGCATTTCATCTTCATACACTGGTTAAACTACATACT     3350
GCCGAGTTGGCCTGCGTAAAGTAGAAGTATGTGACCAATTTGATGTATGA

TCTATGGATCTTTGAATTGAACAAAAAATGATTTTATTCAGAATAATGAT     3400
AGATACCTAGAAACTTAACTTGTTTTTTACTAAAATAAGTCTTATTACTA

AAATACGATTATATATAAA
TTTATGCTAATATATATTT
```

SEQUENCE LISTING

FIG. 3E

```
MNEEFEGDVPMSDPFLSLVTKLDDIAPFPNNDPLDFDMEHNWQEPGPSQQ      50
PDPSIPGNQHSPPQEYYDIDGQRDVSTLHSLLNHNNDDFFSMRFSPPNFD      100
LGGGRGPSLAATQQLSGEGPASMLNPLQTSPPSGGYPPADAYRPLSLAQQ      150
LAAPAMTPHQAASLFVNTNGIDQKNFTHAMLSSPHHTSMTSQPYTEAMGH      200
INGYMSPYDQAQGPSGPSYYSQHHQSPPPHHHHHHPMPKIHENPEQVASP      250
SIEDAPETKPTHLVEPQSPKSPQNMKEELLRLLVNMSPSEVERLKNKKSG      300
ACSATNGPSRSKEKAAKIVIQETAEGDEDEDDEDSDSGETMSQGTTIIVR      350
RPKTERRTAHNLIEKKYRCSINDRIQQLKVLLCGDEAKLSKSATLRRAIE      400
HIEEVEHENQVLKHHVEQMRKTLQNNRLPYPEPIQYTEYSARSPVESSPS      450
PPRNERKRSRMSTTTPMKNGTRDGSSKVTLFAMLLAVLIFNPIGLLAGSA      500
IFSKAAAEAPIASPFEHGRVIDDPDGTSTRTLFWEGSIINMSYVWVFNIL      550
MIIYVVVKLLIHGDPVQDFMSVSWQTFVTTREKARAELNSGNLKDAQRKF      600
CECLATLDRSLPSPGVDSVFSVGWECVRHLLNWLWIGRYIARRRRSTTKP      650
VSVVCRSHAQTAVLYHEIHQLHLMGITGNFEDTYEPSALTGLFMSLCAVN      700
LAEAAGASNDGLPRAVMAQIYISASIQCRLALPNLLAPFFSGYFLRRARR      750
HVRRAPEHSVSHLLWIFHPATRKFMSDAKRLEHVLSSKQKQLRFGSFVED      800
EQLSPLARIRTTLKVYLLSKLVQELVGGDEIFTKNVERILNDNDRLDDEV      850
DVVDVSRLLVTISTQCAAILTNEKDESAKFGTWISRNGDACCTWWTHVLT      900
CGIYWRSNKNELARQHYSLIRNCPPKILTDNLGLAVGHALCARKICIDDR      950
DSPKVSQYVCIHTKKSLESLRLFSTSSRASGVVSGIQEGTRRMAYEWIMN     1000
SLLDAWRSNLFASKPYWTQSFKGQSTFSTLYQEAYNHYAIINGTRGDCWR     1050
LFVYELTCRMLNGANPQATWSGXRRVRSTKMDAVRGRVSMRRSAQPDAFH     1100
LHTLVKLHTSMDL
```

FIG. 4

```
CGGCACGAGGATTAATGCTGATTTCTGGTCTGGACTACACAGCATTGCTG     50
GCCGTGCTCCTAATTACGACTAAAGACCAGACCTGATGTGTCGTAACGAC

GTATAAGGAGTCGGGACCAGAGGAGTAAGATTTCGGGAAGGAATCCCGTC    100
CATATTCCTCAGCCCTGGTCTCCTCATTCTAAAGCCCTTCCTTAGGGCAG

CGGTAGGGACTACTAGCATTCGCAAGTGACGTCCAGCAACCGGAGGACCC    150
GCCATCCCTGATGATCGTAAGCGTTCACTGCAGGTCGTTGGCCTCCTGGG

CCAACTGTAGAATCCGCATCACCATCCTAATCCCAACAAACCAATGACAT    200
GGTTGACATCTTAGGCGTAGTGGTAGGATTAGGGTTGTTTGGTTACTGTA

CTTGAGACCTCACCAGCCATGGATCCCTTCGTGTTCTTCATAGTACTGGC    250
GAACTCTGGAGTGGTCGGTACCTAGGGAAGCACAAGAAGTATCATGACCG

ATCGCTTTATGGCGTTCTTTACTTTTTCGACCGCTTCTTCAAGAGTTGCA    300
TAGCGAAATACCGCAAGAAATGAAAAAGCTGGCGAAGAAGTTCTCAACGT

TGCACTACCCGTACGATGCCTTCCTCAAGAACACCGGGCTGAGTATAAAT    350
ACGTGATGGGCATGCTACGGAAGGAGTTCTTGTGGCCCGACTCATATTTA

TTCATGAGCCTCCACTGGCACACGAGTGCCTTTAACAGGACCCTCCTACG    400
AAGTACTCGGAGGTGACCGTGTGCTCACGGAAATTGTCCTGGGAGGATGC

CTGGGGATCTGCCGGTAACAGCTGCACCCGGAGAGTAATGATCACCAGCT    450
GACCCCTAGACGGCCATTGTCGACGTGGGCCTCTCATTACTAGTGGTCGA

TTAATGTAGGAGTCCTGGTCACCTTTTCTCTGCTCCCGATCGGTCTGATC    500
AATTACATCCTCAGGACCAGTGGAAAAGAGACGAGGGCTAGCCAGACTAG

CTGCTCATTGCCACTATCTTCAGCAGTGGTGAACAAGACAGCTCTTCGTC    550
GACGAGTAACGGTGATAGAAGTCGTCACCACTTGTTCTGTCGAGAAGCAG

TGTATCCTCGCCCGTTGGAGTCCCTGTGCAGCTGGAAATTCTACTGCCCG    600
ACATAGGAGCGGGCAACCTCAGGGACACGTCGACCTTTAAGATGACGGGC

GCGTCAACTTGCCGTTGGAGGAGATCGGATACTACATCACAACCCTTGTG    650
CGCAGTTGAACGGCAACCTCCTCTAGCCTATGATGTAGTGTTGGGAACAC

CTCTGCTTGGTGGTGCACGAGATGGGACACGCCCTGGCCGCTGTGATGGA    700
GAGACGAACCACCACGTGCTCTACCCTGTGCGGGACCGGCGACACTACCT

GGATGTGCCTGTCACCGGGTTTGGAATAAAGTTCATCTTCTGCCTGCCGT    750
CCTACACGGACAGTGGCCCAAACCTTATTTCAAGTAGAAGACGGACGGCA

TAGCATACACGGAGCTCTCCCACGACCACTTAAACAGTCTACGTTGGTTC    800
ATCGTATGTGCCTCGAGAGGGTGCTGGTGAATTTGTCAGATGCAACCAAG
```

FIG. 5A

```
CGCAAGCTACGTGTTCTGTGCGCTGGAATCTGGCATAATTTTGTGTTCGC     850
GCGTTCGATGCACAAGACACGCGACCTTAGACCGTATTAAAACACAAGCG

TGGCGTGTGCTATCTCTTAATCTCAACGGTGGGAATCACTATGTCACCTT     900
ACCGCACACGATAGAGAATTAGAGTTGCCACCCTTAGTGATACAGTGGAA

TGTACGCTTACAACCAACACGTAGTGGTCACTGAACTAACAAGGAAATCC     950
ACATGCGAATGTTGGTTGTGCATCACCAGTGACTTGATTGTTCCTTTAGG

CCGCTGAGGGGAGAGCGCGGCTTGCAAGTGGACAATCAAATAACCCAAGT    1000
GGCGACTCCCCTCTCGCGCCGAACGTTCACCTGTTAGTTTATTGGGTTCA

AAACGGCTGCCCAGTAAACAGCGAGGAGAGTTGGGTGACATGCCTGCAGA    1050
TTTGCCGACGGGTCATTTGTCGCTCCTCTCAACCCACTGTACGGACGTCT

ACTCTCTGAAGCTCAAGCCGGGCTACTGTGTGAGTGCGGACTTCGTGCAG    1100
TGAGAGACTTCGAGTTCGGCCCGATGACACACTCACGCCTGAAGCACGTC

CTTAACGACGAAAGCAGCGCCATCTCACATCATAGCATTGATGGTCAGCT    1150
GAATTGCTGCTTTCGTCGCGGTAGAGTGTAGTATCGTAACTACCAGTCGA

ACAGTGCTGTGATGAACTAAATCCGAACGTAAGCTGCTTCGAGGTGGTGG    1200
TGTCACGACACTACTTGATTTAGGCTTGCATTCGACGAAGCTCCACCACC

AGGACGCAAATGGAGATGTGCCGGTGGAGCTGCCGCAGCATGTATGTCTC    1250
TCCTGCGTTTACCTCTACACGGCCACCTCGACGGCGTCGTACATACAGAG

AATGTGCGCCGCACTTTGGAGGAGGTCTCCGAGCACTGCTCGTCCGGAGT    1300
TTACACGCGGCGTGAAACCTCCTCCAGAGGCTCGTGACGAGCAGGCCTCA

TTGCAACGAGGGATTCTGCCTACGACCGCTTATACGAAATATCACTGCCA    1350
AACGTTGCTCCCTAAGACGGATGCTGGCGAATATGCTTTATAGTGACGGT

TAATGACGTTCAAGCGACAGAATTTTCGCGGAGAGAAGCTGCCGCCGGTG    1400
ATTACTGCAAGTTCGCTGTCTTAAAAGCGCCTCTCTTCGACGGCGGCCAC

ATCTATGTGGGCCATCCATGGGATGTCACTCGAACTGTGGAGGTATCCGC    1450
TAGATACACCCGGTAGGTACCCTACAGTGAGCTTGACACCTCCATAGGCG

CTTTGTGCCGAGATATAGCTTATTAAAGGCAGCCTGGCCGGATGCCTGGC    1500
GAAACACGGCTCTATATCGAATAATTTCCGTCGGACCGGCCTACGGACCG

TGCTGCTCCTCAAGTATAACGTGGTCTTCAGCATAGGATTGGCGTTGATC    1550
ACGACGAGGAGTTCATATTGCACCAGAAGTCGTATCCTAACCGCAACTAG

AATGCCATTCCCTGCTTTGGTTTCGATGGCGCCCACATTACCAGCACCGT    1600
TTACGGTAAGGGACGAAACCAAAGCTACCGCGGGTGTAATGGTCGTGGCA
```

FIG. 5B

```
GATACACAGCTTCTTGGTGGGCAGAGTGGATCAGCATGCCAAGAGAGATA    1650
CTATGTGTCGAAGAACCACCCGTCTCACCTAGTCGTACGGTTCTCTCTAT

TCATCTCGTTGATAATCACCAGCGTGGGTTCCCTTCTCTTTGCACTGGCC    1700
AGTAGAGCAACTATTAGTGGTCGCACCCAAGGGAAGAGAAACGTGACCGG

CTGCTTAAGGTGGCCTGGTTGAGTTTTCTGCGACCCCTGCTTTAAGAACT    1750
GACGAATTCCACCGGACCAACTCAAAAGACGCTGGGGACGAAATTCTTGA

GAAATGGAAAACTGAAATGGATCCTGGGAGTTCAACTCCCTGCAAAGACG    1800
CTTTACCTTTTGACTTTACCTAGGACCCTCAAGTTGAGGGACGTTTCTGC

CTAGACTGCTATTTCACCTTCACGAAACACACAAAAACACAGCGAATTGT    1850
GATCTGACGATAAAGTGGAAGTGCTTTGTGTGTTTTTGTGTCGCTTAACA

AGCACCTCAAAGATTCGATAGCTTTTTGTCATAGTCCTTAGTCTTAACTC    1900
TCGTGGAGTTTCTAAGCTATCGAAAAACAGTATCAGGAATCAGAATTGAG

GTATTTATTTTCGTACGGTTGTCGAGCTCAAAAATAAAATCAAATTAAGC    1950
CATAAATAAAAGCATGCCAACAGCTCGAGTTTTTATTTTAGTTTAATTCG

TAAAAAAAAAAAAAAAAAAAC
ATTTTTTTTTTTTTTTTTTTG
```

Sat Mar 04 2006 02:42:45

FIG. 5C

```
MDPFVFFIVLASLYGVLYFFDRFFKSCMHYPYDAFLKNTGLSINFMSLHW    50
HTSAFNRTLLRWGSAGNSCTRRVMITSFNVGVLVTFSLLPIGLILLIATI    100
FSSGEQDSSSSVSSPVGVPVQLEILLPGVNLPLEEIGYYITTLVLCLVVH    150
EMGHALAAVMEDVPVTGFGIKFIFCLPLAYTELSHDHLNSLRWFRKLRVL    200
CAGIWHNFVFAGVCYLLISTVGITMSPLYAYNQHVVVTELTRKSPLRGER    250
GLQVDNQITQVNGCPVNSEESWVTCLQNSLKLKPGYCVSADFVQLNDESS    300
AISHHSIDGQLQCCDELNPNVSCFEVVEDANGDVPVELPQHVCLNVRRTL    350
EEVSEHCSSGVCNEGFCLRPLIRNITAIMTFKRQNFRGEKLPPVIYVGHP    400
WDVTRTVEVSAFVPRYSLLKAAWPDAWLLLLKYNVVFSIGLALINAIPCF    450
GFDGAHITSTVIHSFLVGRVDQHAKRDIISLIITSVGSLLFALALLKVAW    500
LSFLRPLL
```

FIG. 6

```
GTGTGCCTGACTGTTTTGTAGGTGTAAGGAGGGGCGTGGCCAAATAGTTT     50
CACACGGACTGACAAAACATCCACATTCCTCCCCGCACCGGTTTATCAAA

TTGGTATACGGATAGAATTTGGATGAAAAATAAAACGAAATCAAAACATT    100
AACCATATGCCTATCTTAAACCTACTTTTTATTTTGCTTTAGTTTTGTAA

TTTCAAAAGCGTGGAAGTTTTGGCCGGCTTGTGGGCATGGCAAAACGTTT    150
AAAGTTTTCGCACCTTCAAAACCGGCCGAACACCCGTACCGTTTTGCAAA

TTTGGCTATCCGTTAATCAACATACCGTTGCCCGGGACAATACCCACCAA    200
AAACCGATAGGCAATTAGTTGTATGGCAACGGGCCCTGTTATGGGTGGTT

GATCGTTGTACCCTACGAAACTGGATCCGGATCGCTGTCATGGCACTCTC    250
CTAGCAACATGGGATGCTTTGACCTAGGCCTAGCGACAGTACCGTGAGAG

TTAATACATCCTCGACTACACCGCAGGAACCGCACCCTTCCGGCGAACCC    300
AATTATGTAGGAGCTGATGTGGCGTCCTTGGCGTGGGAAGGCCGCTTGGG

TGGCCCCCCGAACCACAGGTACTCAATAGCAGTACCACGGACCGCAGCCC    350
ACCGGGGGGCTTGGTGTCCATGAGTTATCGTCATGGTGCCTGGCGTCGGG

GCCTCCCCTTCTGCCCTGGGCGCAGAGCAGCCCCGCCTTTTTCTACGTCC    400
CGGAGGGGAAGACGGGACCCGCGTCTCGTCGGGGCGGAAAAAGATGCAGG

AGCAGATTACTCTGCGAACCAGTGTTCTCCCGTGGACGGAGGGAATGCAG    450
TCGTCTAATGAGACGCTTGGTCACAAGAGGGCACCTGCCTCCCTTACGTC

CTTATGGATGCGTTTCGTGCGCCGCTACACGAAGTTTTTAAATTGCTTGA    500
GAATACCTACGCAAAGCACGCGGCGATGTGCTTCAAAAATTTAACGAACT

AATTGTGCGCAATCACCAGAGCAGCGAAAACAAACGTACCCTGGAGCACA    550
TTAACACGCGTTAGTGGTCTCGTCGCTTTTGTTTGCATGGGACCTCGTGT

ACTGCCTACATGTAGACAACGTAAAGCGCGGAACACACGGGCAGCTGGAC    600
TGACGGATGTACATCTGTTGCATTTCGCGCCTTGTGTGCCCGTCGACCTG

CAGATCTTTCCGGAGTATGGCTGCCTGCTGCTCTCGCCCGCCAACCTGTG    650
GTCTAGAAAGGCCTCATACCGACGGACGACGAGAGCGGGCGGTTGGACAC

GACGCAGAACTCTCAGAACTTTACTCGGGACACAAACATCCTGAACACGA    700
CTGCGTCTTGAGAGTCTTGAAATGAGCCCTGTGTTTGTAGGACTTGTGCT

TATTTCAGTACCATAACCTACAGAAATCAAAAGTTTCCGCGGCGGAAATG    750
ATAAAGTCATGGTATTGGATGTCTTTAGTTTTCAAAGGCGCCGCCTTTAC

CTGTTTGGATTACCCATGCAGGACACTGGATTCAAGCGCTATCCATTGCG    800
GACAAACCTAATGGGTACGTCCTGTGACCTAAGTTCGCGATAGGTAACGC
```

FIG. 7A

```
CGCTCGGTCGCGTATTATACAGTATGCCTTGACGTTATTCCTCAAGCACA     850
GCGAGCCAGCGCATAATATGTCATACGGAACTGCAATAAGGAGTTCGTGT

ACGATATGGAGTATCTGGACACTCTAAAGGAAAAGCTGCTGCGACACTAT     900
TGCTATACCTCATAGACCTGTGAGATTTCCTTTTCGACGACGCTGTGATA

CCCCCACTCCCGTTGGCTAGTGCGTCGGCTGAAGAGCCGACGACCATAAC     950
GGGGGTGAGGGCAACCGATCACGCAGCCGACTTCTCGGCTGCTGGTATTG

TTACATCTTTTATCCAGGAGAGTACAGGATGTGGGAGCTGGTGCCTTACA    1000
AATGTAGAAAATAGGTCCTCTCATGTCCTACACCCTCGACCACGGAATGT

CAGTGGCCTTTATGTTGGTGTTTGCTTATGTGTACTTCTCTGTTCGAAAA    1050
GTCACCGGAAATACAACCACAAACGAATACACATGAAGAGACAAGCTTTT

ATCGATGTATTTCGTTCCCGCTTTTTGCTGGCCTTATGTAGCGTAATCAC    1100
TAGCTACATAAAGCAAGGGCGAAAAACGACCGGAATACATCGCATTAGTG

CACAGCCGGGAGCTTGGCCATGTCCCTTGGCTTGTGTTTCTTCTTTGGCC    1150
GTGTCGGCCCTCGAACCGGTACAGGGAACCGAACACAAAGAAGAAACCGG

TGACAATTTCGCTGCAGTCAAAGGACATTTTCCCCTACCTTGTAATCCTT    1200
ACTGTTAAAGCGACGTCAGTTTCCTGTAAAAGGGGATGGAACATTAGGAA

GTGGGATTGGAAAATAGCTTGGTGATCACAAAGAGCGTAGTCTCAATGGA    1250
CACCCTAACCTTTTATCGAACCACTAGTGTTTCTCGCATCAGAGTTACCT

CGAGACATTCGACGTGAAGATCCGCGTGGCGCAGGCTCTTAGCAAGGAGG    1300
GCTCTGTAAGCTGCACTTCTAGGCGCACCGCGTCCGAGAATCGTTCCTCC

GTTGGCATATATCCAAGACTCTTTTGACGGAGATAACAATTTTGACAATT    1350
CAACCGTATATAGGTTCTGAGAAAACTGCCTCTATTGTTAAAACTGTTAA

GGTCTTGCTACTTTCGTGCCCGTCATCCAGGAGTTTTGTATCTTTGCCAT    1400
CCAGAACGATGAAAGCACGGGCAGTAGGTCCTCAAAACATAGAAACGGTA

AGTCGGCTTGCTTTCCGATTTTATGCTACAGATGCTGCTCTTCTCAACAA    1450
TCAGCCGAACGAAAGGCTAAAATACGATGTCTACGACGAGAAGAGTTGTT

TACTGGCCATGAACATTAAGCGGACCGAGTATACGGCGGAGGCCAAGCAC    1500
ATGACCGGTACTTGTAATTCGCCTGGCTCATATGCCGCCTCCGGTTCGTG

CTTCCTAAGATGTTGCTGAGCTGCACCCAAGGGGCTGGTCGACAGGATTT    1550
GAAGGATTCTACAACGACTCGACGTGGGTTCCCCGACCAGCTGTCCTAAA

CCGATTTTTCGGGGCCGCCCCAGCACTGCCACCGTTTGTCCCTGGCACAT    1600
GGCTAAAAAGCCCCGGCGGGGTCGTGACGGTGGCAAACAGGGACCGTGTA
```

FIG. 7B

```
TTCAGCGTTCTCAGTCGCATCCAAAACTGTGTTTTGCTGATCCCGCATCT    1650
AAGTCGCAAGAGTCAGCGTAGGTTTTGACACAAAACGACTAGGGCGTAGA

GTTAGCGATCGTACAAGCTTGGTTAATGGACACTCGTCGCCGGAGCAACG    1700
CAATCGCTAGCATGTTCGAACCAATTACCTGTGAGCAGCGGCCTCGTTGC

AATACCCAAACGCATAAAGATTGTAAATTTCTGGGCGCGGACTCGCTTTT    1750
TTATGGGTTTGCGTATTTCTAACATTTAAAGACCCGCGCCTGAGCGAAAA

TTCAGCGTGCCTTCATGATCTGGATGATTGTGTGGATATGCTCTATAGTT    1800
AAGTCGCACGGAAGTACTAGACCTACTAACACACCTATACGAGATATCAA

TATAATTCGGGATATCTGGAGCAGTTGTTTAGCATGCAGAGCAACGGCAC    1850
ATATTAAGCCCTATAGACCTCGTCAACAAATCGTACGTCTCGTTGCCGTG

AATGACGGCAACCCTTGAACTTCAACGGCGACTACAGGCGGGTCGGGGAG    1900
TTACTGCCGTTGGGAACTTGAAGTTGCCGCTGATGTCCGCCCAGCCCCTC

CAGTCAGCAGTTTTTTCGAGGGATGGCAAGCGGACGGGCAGCGTGCCACG    1950
GTCAGTCGTCAAAAAAGCTCCCTACCGTTCGCCTGCCCGTCGCACGGTGC

AGTGCGCCAAGCGGAAGCGGCTTTTCTACGCCAATAAAAGCTCCTCTAGC    2000
TCACGCGGTTCGCCTTCGCCGAAAAGATGCGGTTATTTTCGAGGAGATCG

GATCGATATAAACGAAACGGCCGAGGAAATGATGAGACTTCGATATCCCA    2050
CTAGCTATATTTGCTTTGCCGGCTCCTTTACTACTCTGAAGCTATAGGGT

GCTTCGACCTAAACTATTTCCTTTCAAACTTCCACTGGTCCACGATTATG    2100
CGAAGCTGGATTTGATAAAGGAAAGTTTGAAGGTGACCAGGTGCTAATAC

AAACAGTACAACATCTCACTAAGTGGGCACTACGTTACCCTGCTACCGAC    2150
TTTGTCATGTTGTAGAGTGATTCACCCGTGATGCAATGGGACGATGGCTG

CATTCGCCTTAGTCATGCCATCGCTCCGGAGCTAGCCACTCTGTTGCGGA    2200
GTAAGCGGAATCAGTACGGTAGCGAGGCCTCGATCGGTGAGACAACGCCT

ATCCGCAGGAGCAGCTGCAACAAAATTTTCAATGGAAGGCCCTAGCCGCT    2250
TAGGCGTCCTCGTCGACGTTGTTTTAAAAGTTACCTTCCGGGATCGGCGA

GCACTCGATCCGCTGGACTTTAACGATGACGACGTGCGCCGTGAGTCTCC    2300
CGTGAGCTAGGCGACCTGAAATTGCTACTGCTGCACGCGGCACTCAGAGG

GATGGTAATGGCAGAGGGGTTGCCTCTGGTTCCCAAGAGCCCCATGGAAA    2350
CTACCATTACCGTCTCCCCAACGGAGACCAAGGGTTCTCGGGGTACCTTT

TATTTTTCGCCATCCTCTTGTGCTGCATCAGCATCTTCGTGCTTTGCTAC    2400
ATAAAAAGCGGTAGGAGAACACGACGTAGTCGTAGAAGCACGAAACGATG
```

FIG. 7C

ACGATGGTGGTTTTCTACCGCTGCATATGTACCAGGAACTATGCCGAGTG 2450
TGCTACCACCAAAAGATGGCGACGTATACATGGTCCTTGATACGGCTCAC

GCGCTCCAGTTGGCACGAATCTGAGGCACCGTACAAGCAGACTGAGCAAA 2500
CGCGAGGTCAACCGTGCTTAGACTCCGTGGCATGTTCGTCTGACTCGTTT

TCCTGGAGGGAGTTCCAACGCAAATCGCCGGACACAAACATCGCATTGAA 2550
AGGACCTCCCTCAAGGTTGCGTTTAGCGGCCTGTGTTTGTAGCGTAACTT

TGCCTGGTGTCTGACGGCGCCTACATAATCAGCTGCTGCCTTAAAGGCCA 2600
ACGGACCACAGACTGCCGCGGATGTATTAGTCGACGACGGAATTTCCGGT

AATCCGAGTGTGGGATGCACGCAGTGGCGAGCAGCTAACCAGCATCTCCC 2650
TTAGGCTCACACCCTACGTGCGTCACCGCTCGTCGATTGGTCGTAGAGGG

GATCCGATATTCAGATCTCTCAGCAGCGGACGGATGGGCAGACGCTGGTA 2700
CTAGGCTATAAGTCTAGAGAGTCGTCGCCTGCCTACCCGTCTGCGACCAT

CGAAAGCTGGCCGTGTCACCGGTCTGGTGCCTTGACTACTTCGATAATCT 2750
GCTTTCGACCGGCACAGTGGCCAGACCACGGAACTGATGAAGCTATTAGA

AATCGCAGTAGGCTGCGCCAACGGCCGCGTAGAATTGTGGGAATCCCCTG 2800
TTAGCGTCATCCGACGCGGTTGCCGGCGCATCTTAACACCCTTAGGGGAC

CGGGATTGCTTAAGTGTGCATACCAGGAAGACGCGAAGAGAAACCAGGGT 2850
GCCCTAACGAATTCACACGTATGGTCCTTCTGCGCTTCTCTTTGGTCCCA

ATAACCCACATCCACCTGAACGGCGATCGAGTGATTGTGGCGCGTCTTAA 2900
TATTGGGTGTAGGTGGACTTGCCGCTAGCTCACTAACACCGCGCAGAATT

TGGCCGACTAGATTTTTACCGCTTAGAGACGTACTACAAGGGGAAGCAAA 2950
ACCGGCTGATCTAAAAATGGCGAATCTCTGCATGATGTTCCCCTTCGTTT

TCGACTGGGGTTTTACCTCGGCTTACAGGAGAACTCATGTTCGAACTGGA 3000
AGCTGACCCCAAAATGGAGCCGAATGTCCTCTTGAGTACAAGCTTGACCT

TCCACTGGAAGCCTGGGATTAATGTTGCAGCAGCAGCGCTGTCAGCAAGA 3050
AGGTGACCTTCGGACCCTAATTACAACGTCGTCGTCGCGACAGTCGTTCT

AGCATCCCAGAAGACCACCAAGGAGGAAATGAAAATCACATTGGAGGGTG 3100
TCGTAGGGTCTTCTGGTGGTTCCTCCTTTACTTTTAGTGTAACCTCCCAC

TAAGACTAGCCCATCAGCAGCCAATCACATGCATGCAGGTCGTTAACGAC 3150
ATTCTGATCGGGTAGTCGTCGGTTAGTGTACGTACGTCCAGCAATTGCTG

ATGGTTTTCACTGGCAGCCAGGATCACACCCTCAAGGTGTATTGCCTCAA 3200
TACCAAAAGTGACCGTCGGTCCTAGTGTGGGAGTTCCACATAACGGAGTT

FIG. 7D

```
TAAGTCGGATGTTGAGTATACGCTCCACGGTCACTGTGGGCCTGTAACCT      3250
ATTCAGCCTACAACTCATATGCGAGGTGCCAGTGACACCCGGACATTGGA

GTCTCTTTGTGGATCGCTGGCAACCTGGCACAGGGGGGTCTGGGTCCCAG      3300
CAGAGAAACACCTAGCGACCGTTGGACCGTGTCCCCCCAGACCCAGGGTC

GACGGCCTGCTCTGCGTATGGGATCTGTTCACGGGAGCCTGCATGTATAA      3350
CTGCCGGACGAGACGCATACCCTAGACAAGTGCCCTCGGACGTACATATT

TATACAAGCTCACGACGGAGCCGTCAGCTGCCTGGCCTGTGCGCCCAGTT      3400
ATATGTTCGAGTGCTGCCTCGGCAGTCGACGGACCGGACACGCGGGTCAA

ACGTAATCTCGCTAGGCACGGACGAGAGGATTTGCGTATGGGAACGATTT      3450
TGCATTAGAGCGATCCGTGCCTGCTCTCCTAAACGCATACCCTTGCTAAA

CAGGGAAACCTGTTGACTACCATCAACATCTCAAACGCATACTCGAGCCT      3500
GTCCCTTTGGACAACTGATGGTAGTTGTAGAGTTTGCGTATGAGCTCGGA

ACTGATGCTAACACCGTCACTATTGGTTACGAGCAAAATGGGTAAGGCCT      3550
TGACTACGATTGTGGCAGTGATAACCAATGCTCGTTTTACCCATTCCGGA

CATTCTTGATTGCCAATATAAGAGGGACAGTAAATAATAAATTTAATTCC      3600
GTAAGAACTAACGGTTATATTCTCCCTGTCATTTATTATTTAAATTAAGG

AACACAGGATCTCTTATTGTGTGGGATGTGCGCACTGGGCAGCCGGCTCG      3650
TTGTGTCCTAGAGAATAACACACCCTACACGCGTGACCCGTCGGCCGAGC

CGAGGTCAAACTGGACTTTGCAAACCTGCAGCTCTGTCCCAAAATAATGA      3700
GCTCCAGTTTGACCTGAAACGTTTGGACGTCGAGACAGGGTTTTATTACT

TGCTTGCCTGCGATTCGGTAGTTTGCGACTACGGAAATGAGATCCGCGTC      3750
ACGAACGGACGCTAAGCCATCAAACGCTGATGCCTTTACTCTAGGCGCAG

GTCCGCTTTCCTATCGTGGCAGACAAGTGCCATTAAAGCGCAAAATTTTA      3800
CAGGCGAAAGGATAGCACCGTCTGTTCACGGTAATTTCGCGTTTTAAAAT

ATTTAGCGTGGTTCGCTAGCACCTAGGAATAAGTTGACTTAAGGCTTTAA      3850
TAAATCGCACCAAGCGATCGTGGATCCTTATTCAACTGAATTCCGAAATT

AACGCCTGGAAGTCATTGACGCATTCACTATTTTATATAAATATATACAC      3900
TTGCGGACCTTCAGTAACTGCGTAAGTGATAAAATATATTTATATATGTG

TATTAGGGTCCGCAGCAACTTACGGTTTTAACACAAGCTGTACGTATCTC      3950
ATAATCCCAGGCGTCGTTGAATGCCAAAATTGTGTTCGACATGCATAGAG

ATCTCTAGAATTTTGTGTTAGTTTGTGGACACTAAGTGTAACAGCTACGC      4000
TAGAGATCTTAAAACACAATCAAACACCTGTGATTCACATTGTCGATGCG
```

FIG. 7E

```
TCCGGTAGGTTAAGGAACTAAACTAAATGAATCAGATATATACACATATA    4050
AGGCCATCCAATTCCTTGATTTGATTTACTTAGTCTATATATGTGTATAT

TTTTCGCGTAATTATATAAACTACATAGTGTCTTAAAGCGCCTCAGCCTA    4100
AAAAGCGCATTAATATATTTGATGTATCACAGAATTTCGCGGAGTCGGAT

ATATAAAATGACTAAATGTTAAAATAAA
TATATTTTACTGATTTACAATTTTATTT
```

FIG. 7F

MKNKTKSKHFSKAWKFWPACGHGKTFFGYPLINIPLPGTIPTKIVVPYET 50

GSGSLSWHSLNTSSTTPQEPHPSGEPWPPEPQVLNSSTTDRSPPPLLPWA 100

QSSPAFFYVQQITLRTSVLPWTEGMQLMDAFRAPLHEVFKLLEIVRNHQS 150

SENKRTLEHNCLHVDNVKRGTHGQLDQIFPEYGCLLLSPANLWTQNSQNF 200

TRDTNILNTIFQYHNLQKSKVSAAEMLFGLPMQDTGFKRYPLRARSRIIQ 250

YALTLFLKHNDMEYLDTLKEKLLRHYPPLPLASASAEEPTTITYIFYPGE 300

YRMWELVPYTVAFMLVFAYVYFSVRKIDVFRSRFLLALCSVITTAGSLAM 350

SLGLCFFFGLTISLQSKDIFPYLVILVGLENSLVITKSVVSMDETFDVKI 400

RVAQALSKEGWHISKTLLTEITILTIGLATFVPVIQEFCIFAIVGLLSDF 450

MLQMLLFSTILAMNIKRTEYTAEAKHLPKMLLSCTQGAGRQDFRFFGAAP 500

ALPPFVPGTFQRSQSHPKLCFADPASVSDRTSLVNGHSSPEQRIPKRIKI 550

VNFWARTRFFQRAFMIWMIVWICSIVYNSGYLEQLFSMQSNGTMTATLEL 600

QRRLQAGRGAVSSFFEGWQADGQRATSAPSGSGFSTPIKAPLAIDINETA 650

EEMMRLRYPSFDLNYFLSNFHWSTIMKQYNISLSGHYVTLLPTIRLSHAI 700

APELATLLRNPQEQLQQNFQWKALAAALDPLDFNDDDVRRESPMVMAEGL 750

PLVPKSPMEIFFAILLCCISIFVLCYTMVVFYRCICTRNYAEWRSSWHES 800

EAPYKQTEQILEGVPTQIAGHKHRIECLVSDGAYIISCCLKGQIRVWDAR 850

SGEQLTSISRSDIQISQQRTDGQTLVRKLAVSPVWCLDYFDNLIAVGCAN 900

GRVELWESPAGLLKCAYQEDAKRNQGITHIHLNGDRVIVARLNGRLDFYR 950

LETYYKGKQIDWGFTSAYRRTHVRTGSTGSLGLMLQQQRCQQEASQKTTK 1000

EEMKITLEGVRLAHQQPITCMQVVNDMVFTGSQDHTLKVYCLNKSDVEYT 1050

LHGHCGPVTCLFVDRWQPGTGGSGSQDGLLCVWDLFTGACMYNIQAHDGA 1100

VSCLACAPSYVISLGTDERICVWERFQGNLLTTINISNAYSSLLMLTPSL 1150

LVTSKMGKASFLIANIRGTVNNKFNSNTGSLIVWDVRTGQPAREVKLDFA 1200

NLQLCPKIMMLACDSVVCDYGNEIRVVRFPIVADKCH

FIG. 8

```
GTTTATTAAGCTGCAAATATACTCGTGAAAAAAATCAAAACAACCATGAA      50
CAAATAATTCGACGTTTATATGAGCACTTTTTTTAGTTTTGTTGGTACTT

CAACAAGTGTTGCAACTATTACTAACTAGTCGCTAGTTTAAAGCAAAGTG     100
GTTGTTCACAACGTTGATAATGATTGATCAGCGATCAAATTTCGTTTCAC

CGTTGACATTAACCAGTTATGGAAAAACAAAAGCACACGTGAACTAAGAA     150
GCAACTGTAATTGGTCAATACCTTTTTGTTTTCGTGTGCACTTGATTCTT

AACAGATAGAAGGTGGTAAAGCATTCGCAATGGACACGACACTGATGAAC     200
TTGTCTATCTTCCACCATTTCGTAAGCGTTACCTGTGCTGTGACTACTTG

TTAATAGACGCTCCGCTGGACGAGTCCATGGATTTGTTCAAAGCGGAGGA     250
AATTATCTGCGAGGCGACCTGCTCAGGTACCTAAACAAGTTTCGCCTCCT

TGTCTTCGAACCGTTCGACGCCGACCTGCACTCGGACATGCTGGACATCA     300
ACAGAAGCTTGGCAAGCTGCGGCTGGACGTGAGCCTGTACGACCTGTAGT

TCCTCAACGACATGGACCTGGCGCCGACGCAGATGTACAACATGCTGCTG     350
AGGAGTTGCTGTACCTGGACCGCGGCTGCGTCTACATGTTGTACGACGAC

GACGAGCCTCGAACGCATACCCAGCAGACGCAGTCCGTGGATCAGCAGCC     400
CTGCTCGGAGCTTGCGTATGGGTCGTCTGCGTCAGGCACCTAGTCGTCGG

GCAATCCGTCGAGCAACAGCCGCACGTGAAAAGCGAGCACTCTTCGCCAG     450
CGTTAGGCAGCTCGTTGTCGGCGTGCACTTTTCGCTCGTGAGAAGCGGTC

TGCACATCAAGGAGGAACTGCATCAGCAGCAACAACAGTCGCCGCTTCTC     500
ACGTGTAGTTCCTCCTTGACGTAGTCGTCGTTGTTGTCAGCGGCGAAGAG

GTCTACAAACCAGATCCCCTCATAGCCACAAGCTACAATTGTCCCCAGCA     550
CAGATGTTTGGTCTAGGGGAGTATCGGTGTTCGATGTTAACAGGGGTCGT

ACAGCCGACGGGCCTTTTGAAGGCCGCCCAACCAACAGCCACCATACATC     600
TGTCGGCTGCCCGGAAAACTTCCGGCGGGTTGGTTGTCGGTGGTATGTAG

ACATGGACGCCCAGCGGATGCCGCCGAACACGGCGGTGTATCCCCCATCT     650
TGTACCTGCGGGTCGCCTACGGCGGCTTGTGCCGCCACATAGGGGGTAGA

CTGGGCAGTAGCTTTGTCTACCAGTCCATGTCCCCGCCCACGTCGCCGGT     700
GACCCGTCATCGAAACAGATGGTCAGGTACAGGGGCGGGTGCAGCGGCCA

GGAGTCTGCGAACCAGAATGTCAATGTCATGCAGCCCGTTGCTGCAACTC     750
CCTCAGACGCTTGGTCTTACAGTTACAGTACGTCGGGCAACGACGTTGAG

CTGCTCCCGCTTCTGCTCCTTTGCCCCAGCAGTCGTATCCGCAACCCTTC     800
GACGAGGGCGAAGACGAGGAAACGGGGTCGTCAGCATAGGCGTTGGGAAG
```

FIG. 9A

```
ATTACGTACAACTCTAAGGCCGGAATGACTTCCGATGAAGCCATGTACTT    850
TAATGCATGTTGAGATTCCGGCCTTACTGAAGGCTACTTCGGTACATGAA

GCTCTTGCAGCCCACGGTAGCCAGTCCAACCCCATCTCCACCTGTGGCTC    900
CGAGAACGTCGGGTGCCATCGGTCAGGTTGGGGTAGAGGTGGACACCGAG

CACCACCGACAAGCACAGGTAGTCGGGCCAGCAAGGTGCGAGTGGCACCA    950
GTGGTGGCTGTTCGTGTCCATCAGCCCGGTCGTTCCACGCTCACCGTGGT

CTGGCTCCGTCACCTGCCGCTATGGAAGTCCAGGGCAAGGTACCTATCAA   1000
GACCGAGGCAGTGGACGGCGATACCTTCAGGTCCCGTTCCATGGATAGTT

CCGGGTTCAACCCAAGGTGAAGGAAGTAAAGCGCTCGGCCCACAACGCCA   1050
GGCCCAAGTTGGGTTCCACTTCCTTCATTTCGCGAGCCGGGTGTTGCGGT

TCGAGCGGCGCTATCGCACCTCAATCAACGACAAGATTAACGAGTTGAAG   1100
AGCTCGCCGCGATAGCGTGGAGTTAGTTGCTGTTCTAATTGCTCAACTTC

AACTTGGTAGTGGGAGAGCAGGCCAAGCTGAACAAGTCCGCAGTGTTGCG   1150
TTGAACCATCACCCTCTCGTCCGGTTCGACTTGTTCAGGCGTCACAACGC

GAAATCCATAGACAAGATTCGGGATCTGCAACGCCAGAATCACGATCTGA   1200
CTTTAGGTATCTGTTCTAAGCCCTAGACGTTGCGGTCTTAGTGCTAGACT

AGGCAGAGTTGCAGCGCCTGCAGAGGGAGCTAATGGCACGCGACGGCTCC   1250
TCCGTCTCAACGTCGCGGACGTCTCCCTCGATTACCGTGCGCTGCCGAGG

AAGGTGAAGGATTTACTTCAGCTGGGCACTCGGCCTGGTAGAGCATCCAA   1300
TTCCACTTCCTAAATGAAGTCGACCCGTGAGCCGGACCATCTCGTAGGTT

GAAGCGCCGCGAGAGCTCGCAGACCTTTACCACGGATGCCGGACTGACGC   1350
CTTCGCGGCGCTCTCGAGCGTCTGGAAATGGTGCCTACGGCCTGACTGCG

CGCCACGCAGCGATGAATCGGATCCTTCGCTCTCGCCCATGCACTCGGAC   1400
GCGGTGCGTCGCTACTTAGCCTAGGAAGCGAGAGCGGGTACGTGAGCCTG

ATCTCGTTGCCGCCATCACCCTATGGTGGATCCACCGCCAGCTGTAGCAG   1450
TAGAGCAACGGCGGTAGTGGGATACCACCTAGGTGGCGGTCGACATCGTC

TGGCAGCAGCAGCAGCAATGAAGAACCACTGGTGGTGCCCAGCTCTATGC   1500
ACCGTCGTCGTCGTCGTTACTTCTTGGTGACCACCACGGGTCGAGATACG

GCGGCATGGCCACCCACTCTCGCCTCGGACTCTGCATGTTTATGTTCGCC   1550
CGCCGTACCGGTGGGTGAGAGCGGAGCCTGAGACGTACAAATACAAGCGG

ATCCTGGCCGTCAATCCCTTCAAGACCTTTCTCCAGCGCGGCCACTATGA   1600
TAGGACCGGCAGTTAGGGAAGTTCTGGAAAGAGGTCGCGCCGGTGATACT
```

FIG. 9B

```
CAGTAATGACGATCTTGGCGACATGAGCGGTCAAAGACGCATTCTCTCTT      1650
GTCATTACTGCTAGAACCGCTGTACTCGCCAGTTTCTGCGTAAGAGAGAA

ACGACGTGGAAGGTGAAGGTTTTGCTGTCTGGCAGCAGAGTTCCTGGATA      1700
TGCTGCACCTTCCACTTCCAAAACGACAGACCGTCGTCTCAAGGACCTAT

TGGCTATTGAACTTCACACTGATGCTTGGATGCTTGGTGAAATTGCTGGT      1750
ACCGATAACTTGAAGTGTGACTACGAACCTACGAACCACTTTAACGACCA

TTACGGTGATCCGCAGCTGGACGCGCAAACGGACGCCTACTGCCAGCACA      1800
AATGCCACTAGGCGTCGACCTGCGCGTTTGCCTGCGGATGACGGTCGTGT

GGCAGCGGGCTGACTTCTATTTTAGCCAAGGACAGTCGTCTCAGGCCTAC      1850
CCGTCGCCCGACTGAAGATAAAATCGGTTCCTGTCAGCAGAGTCCGGATG

GCCGGTTACCTCAACTGTCTGCATATGTTTGGATTAAGTCTACCGGCGTC      1900
CGGCCAATGGAGTTGACAGACGTATACAAACCTAATTCAGATGGCCGCAG

GCGCTTGGAGTGTTACTTGCAGACCACGTGGCAGTTCCTTCGTTTTCTTT      1950
CGCGAACCTCACAATGAACGTCTGGTGCACCGTCAAGGAAGCAAAAGAAA

TCCATCGCCTCTGGCTGGGTCGGGTGCTGTCACGGCGGTCCGGTGGGCTG      2000
AGGTAGCGGAGACCGACCCAGCCCACGACAGTGCCGCCAGGCCACCCGAC

TTTAGCAACGCCGCCAGCAGGAAACAGGCGCTGGCATCTGCACGCGAACT      2050
AAATCGTTGCGGCGGTCGTCCTTTGTCCGCGACCGTAGACGTGCGCTTGA

GGCCCTGCTCTTCAACCGACTGAATCAATTGCAACTGACTGGAAATGGAA      2100
CCGGGACGAGAAGTTGGCTGACTTAGTTAACGTTGACTGACCTTTACCTT

GCCGCGGTGACATGAACGGCATTATGATGGCACTATTCGCAAGCAACATG      2150
CGGCGCCACTGTACTTGCCGTAATACTACCGTGATAAGCGTTCGTTGTAC

GCTGAAGTGGCGCACAATCTACTGACACCGCGCGAGACCATCTGCATCCA      2200
CGACTTCACCGCGTGTTAGATGACTGTGGCGCGCTCTGGTAGACGTAGGT

CGTAATGACAGCGTTGCGAATGAAGCGCAGTGCCCCAAAATGGTTGCAAC      2250
GCATTACTGTCGCAACGCTTACTTCGCGTCACGGGGTTTTACCAACGTTG

AGTTCTTCGCCCGATACTACATGAGCCGGGCTCGTCAAGAGTGCGGTCGC      2300
TCAAGAAGCGGGCTATGATGTACTCGGCCCGAGCAGTTCTCACGCCAGCG

ACTAGGGCCACCGAGCAAACGCAGGAGCTACGTTGGGCATTCACAGCCTA      2350
TGATCCCGGTGGCTCGTTTGCGTCCTCGATGCAACCCGTAAGTGTCGGAT

TGGATATCGCTACTGCGCCACGCACGTCTTCACGTACGATCTGAGCGACT      2400
ACCTATAGCGATGACGCGGTGCGTGCAGAAGTGCATGCTAGACTCGCTGA
```

FIG. 9C

```
CCGGCGAGCAGGATGGATTCTTCACACGTCTTAGGAATCCATGTGATCCC      2450
GGCCGCTCGTCCTACCTAAGAAGTGTGCAGAATCCTTAGGTACACTAGGG

GCTGCCCACGTCATTAAGCAATATCGAGAGCATTTGCTGTTTAAATCCAT      2500
CGACGGGTGCAGTAATTCGTTATAGCTCTCGTAAACGACAAATTTAGGTA

TCAGTGTCTGGTAGGAGCGGGCCACAAATCGGGAGGCCTGCCCACATCTT      2550
AGTCACAGACCATCCTCGCCCGGTGTTTAGCCCTCCGGACGGGTGTAGAA

CTGTCAGCGGAGAGGCGGAACAGTTGCAGCAACAGCAGCACAGCGGCACC      2600
GACAGTCGCCTCTCCGCCTTGTCAACGTCGTTGTCGTCGTGTCGCCGTGG

ATTGTCAGCAATGTTCTTAAGTACACGTCCCTCCTTAAGGACACTCTCTG      2650
TAACAGTCGTTACAAGAATTCATGTGCAGGGAGGAATTCCTGTGAGAGAC

GGCTGATGAGGATGAGCGGGATACAAACGTGGTGTGGTGGGCCGATGTTT      2700
CCGACTACTCCTACTCGCCCTATGTTTGCACCACACCACCCGGCTACAAA

TGGAGACCGCAGTGCACTGGCTCCTTGGTGAAGACACGCTGGCCGAGCAA      2750
ACCTCTGGCGTCACGTGACCGAGGAACCACTTCTGTGCGACCGGCTCGTT

TTGTACGGCAGGATCAAGCAAATGCCCACGCAGCTGCAACAGTGCGGCGA      2800
AACATGCCGTCCTAGTTCGTTTACGGGTGCGTCGACGTTGTCACGCCGCT

AAACGATCATCTGCCCAAGGCGCTGCATGCTGTGCTGCGAGCTAAGATGA      2850
TTTGCTAGTAGACGGGTTCCGCGACGTACGACACGACGCTCGATTCTACT

TCTTACTAAAAAACAATGGCAACGCACTGGACAAAAGTCTCAAGCAATTG      2900
AGAATGATTTTTTGTTACCGTTGCGTGACCTGTTTTCAGAGTTCGTTAAC

GTAAACATCCTCTGCGATGAGTCGAGTGTGGAGCTCCAAGAGTGCTTGAC      2950
CATTTGTAGGAGACGCTACTCAGCTCACACCTCGAGGTTCTCACGAACTG

TGTCAACCGGATCACCGACGCCAAGGGTATAAAGCTGCTTTTCCAGTTGC      3000
ACAGTTGGCCTAGTGGCTGCGGTTCCCATATTTCGACGAAAAGGTCAACG

TTACCTGCGATTGGCTGCTCGAAACTAGGACTGCTCTGTGGGAACTGGAA      3050
AATGGACGCTAACCGACGAGCTTTGATCCTGACGAGACACCCTTGACCTT

CACATGAATATGGAGGACGATGGCTTCTACCAAGTGCCAGGTGAAGTGCT      3100
GTGTACTTATACCTCCTGCTACCGAAGATGGTTCACGGTCCACTTCACGA

CGAGAAGTTCCAGACCGATTTGAACTCGTTGCGCAACATTGTGGAGAATA      3150
GCTCTTCAAGGTCTGGCTAAACTTGAGCAACGCGTTGTAACACCTCTTAT

TACCGAACGCCCAATCGCGCATATATTTGTACGAGGCAGTTTGTCGCCTG      3200
ATGGCTTGCGGGTTAGCGCGTATATAAACATGCTCCGTCAAACAGCGGAC
```

FIG. 9D

```
ATGGCTGGAGCCTCACCGTGTCCAACGCAACAGCTCTTGGACAGGAGTCT    3250
TACCGACCTCGGAGTGGCACAGGTTGCGTTGTCGAGAACCTGTCCTCAGA

GCGATCACGCAACGCCCACTCGTCCATCTTCTGCGGCAGCAAGGATCGGC    3300
CGCTAGTGCGTTGCGGGTGAGCAGGTAGAAGACGCCGTCGTTCCTAGCCG

GGCAGCAGAACTTCGTGGGCGGAGAGCGGGAACGGGCTTCGGCCATGTAC    3350
CCGTCGTCTTGAAGCACCCGCCTCTCGCCCTTGCCCGAAGCCGGTACATG

GTGGCCTGCAAGTATCTCCCGCCTGCGCTGCTCAGCTCCCCGGGTGAACG    3400
CACCGGACGTTCATAGAGGGCGGACGCGACGAGTCGAGGGGCCCACTTGC

TGCTGGCATGTTAGCCGAGGCGGCCAAGACCCTGGAGAAGGTGGGCGACA    3450
ACGACCGTACAATCGGCTCCGCCGGTTCTGGGACCTCTTCCACCCGCTGT

AGCGAAAGCTCAAGGAGTGCTACCAGCTGATGAAGTCGCTGGGCAACGGC    3500
TCGCTTTCGAGTTCCTCACGATGGTCGACTACTTCAGCGACCCGTTGCCG

ATTGGCAGCGTGAAGGCTTAGGATAGTAGTGAAGTACATAATAAGTGGCA    3550
TAACCGTCGCACTTCCGAATCCTATCATCACTTCATGTATTATTCACCGT

CGAACGTGGTGTGGATTTTCAGCAAATGAATACCCGTTTGCTATTCAAAA    3600
GCTTGCACCACACCTAAAAGTCGTTTACTTATGGGCAAACGATAAGTTTT

GAATTACAAATGCCTAGGTCTTTATAATTACGCTATTCCTCTGTTTTCCA    3650
CTTAATGTTTACGGATCCAGAAATATTAATGCGATAAGGAGACAAAAGGT

CGCCCGGTTATGCTTAGATTGTAATTTTAAAATTATTTAATATGGACATT    3700
GCGGGCCAATACGAATCTAACATTAAAATTTTAATAAATTATACCTGTAA

TTATTTGTTTATTATTTACCGTACTTGTTAAACGTATTTATAACAATAAA    3750
AATAAACAAATAATAAATGGCATGAACAATTTGCATAAATATTGTTATTT

TATTTTAACAGATTTAAA
ATAAAATTGTCTAAATTT
```

FIG. 9E

MDTTLMNLIDAPLDESMDLFKAEDVFEPFDADLHSDMLDIILNDMDLAPT 50

QMYNMLLDEPRTHTQQTQSVDQQPQSVEQQPHVKSEHSSPVHIKEELHQQ 100

QQQSPLLVYKPDPLIATSYNCPQQQPTGLLKAAQPTATIHHMDAQRMPPN 150

TAVYPPSLGSSFVYQSMSPPTSPVESANQNVNVMQPVAATPAPASAPLPQ 200

QSYPQPFITYNSKAGMTSDEAMYLLLQPTVASPTPSPPVAPPPTSTGSRA 250

SKVRVAPLAPSPAAMEVQGKVPINRVQPKVKEVKRSAHNAIERRYRTSIN 300

DKINELKNLVVGEQAKLNKSAVLRKSIDKIRDLQRQNHDLKAELQRLQRE 350

LMARDGSKVKDLLQLGTRPGRASKKRRESSQTFTTDAGLTPPRSDESDPS 400

LSPMHSDISLPPSPYGGSTASCSSGSSSSNEEPLVVPSSMRGMATHSRLG 450

LCMFMFAILAVNPFKTFLQRGHYDSNDDLGDMSGQRRILSYDVEGEGFAV 500

WQQSSWIWLLNFTLMLGCLVKLLVYGDPQLDAQTDAYCQHRQRADFYFSQ 550

GQSSQAYAGYLNCLHMFGLSLPASRLECYLQTTWQFLRFLFHRLWLGRVL 600

SRRSGGLFSNAASRKQALASARELALLFNRLNQLQLTGNGSRGDMNGIMM 650

ALFASNMAEVAHNLLTPRETICIHVMTALRMKRSAPKWLQQFFARYYMSR 700

ARQECGRTRATEQTQELRWAFTAYGYRYCATHVFTYDLSDSGEQDGFFTR 750

LRNPCDPAAHVIKQYREHLLFKSIQCLVGAGHKSGGLPTSSVSGEAEQLQ 800

QQQHSGTIVSNVLKYTSLLKDTLWADEDERDTNVVWWADVLETAVHWLLG 850

EDTLAEQLYGRIKQMPTQLQQCGENDHLPKALHAVLRAKMILLKNNGNAL 900

DKSLKQLVNILCDESSVELQECLTVNRITDAKGIKLLFQLLTCDWLLETR 950

TALWELEHMNMEDDGFYQVPGEVLEKFQTDLNSLRNIVENIPNAQSRIYL 1000

YEAVCRLMAGASPCPTQQLLDRSLRSRNAHSSIFCGSKDRRQQNFVGGER 1050

ERASAMYVACKYLPPALLSSPGERAGMLAEAAKTLEKVGDKRKLKECYQL 1100

MKSLGNGIGSVKA

FIG. 10

DECLARATION
AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below at 201 et seq. underneath my name.

I believe I am the original, first and sole inventor if only one name is listed at 201 below, or an original, first and joint inventor if plural names are listed at 201 et seq. below, of the subject matter which is claimed and for which a patent is sought on the invention entitled

ANIMAL MODELS AND METHODS FOR ANALYSIS OF LIPID METABOLISM AND SCREENING OF PHARMACEUTICAL AND PESTICIDAL AGENTS THAT MODULATE LIPID METABOLISM

and for which a patent application:
☒ is attached hereto and includes amendment(s) filed on *(if applicable)*
☐ was filed in the United States on as Application No. *(for declaration not accompanying application)*
with amendment(s) filed on *(if applicable)*
☐ was filed as PCT international Application No. on and was amended under PCT Article 19 on *(if applicable)*

I hereby state that I have reviewed and understand the contents of the above identified application, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

EARLIEST FOREIGN APPLICATION(S), IF ANY, FILED PRIOR TO THE FILING DATE OF THE APPLICATION			
APPLICATION NUMBER	COUNTRY	DATE OF FILING (day, month, year)	PRIORITY CLAIMED
			YES ☐ NO ☐
			YES ☐ NO ☐

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

APPLICATION NUMBER	FILING DATE

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code §112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, §1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application:

APPLICATION SERIAL NO.	FILING DATE	STATUS		
		PATENTED	PENDING	ABANDONED

POWER OF ATTORNEY: As a named inventor, I hereby appoint S. Leslie Misrock (Reg. No. 18872), Harry C. Jones, III (Reg. No. 20280), Berj A. Terzian (Reg. No. 20060), Gerald J. Flintoft (Reg. No. 20823), David Weild, III (Reg. No. 21094), Jonathan A. Marshall (Reg. No. 24614), Barry D. Rein (Reg. No. 22411), Stanton T. Lawrence, III (Reg. No. 25736), Charles E. McKenney (Reg. No. 22795), Philip T. Shannon (Reg. No. 24278), Francis E. Morris (Reg. No. 24615), Charles E. Miller (Reg. No. 24576), Gidon D. Stern (Reg. No. 27469), John J. Lauter, Jr. (Reg. No. 27814), Brian M. Poissant (Reg. No. 28462), Brian D. Coggio (Reg. No. 27624), Rory J. Radding (Reg. No. 28749), Stephen J. Harbulak (Reg. No. 29166), Donald J. Goodell (Reg. No. 19766), James N. Palik (Reg. No. 25510), Thomas E. Friebel (Reg. No. 29258), Laura A. Coruzzi (Reg. No. 30742), Jennifer Gordon (Reg. No. 30753), Jon R. Stark (Reg. No. 30111), Allan A. Fanucci (Reg. No. 30256), Geraldine F. Baldwin (Reg. No. 31232), Victor N. Balancia (Reg. No. 31231), Samuel B. Abrams (Reg. No. 30605), Steven I. Wallach (Reg. No. 35402), Marcia H. Sundeen (Reg. No. 30893), Paul J. Zegger (Reg. No. 33821), Edmond R. Bannon (Reg. No. 32110), Bruce J. Barker (Reg. No. 33291), Adriane M. Antler (Reg. No. 32605), Thomas G. Rowan (Reg. No. 34419), James G. Markey (Reg. No. 31636), Thomas D. Kohler (Reg. No. 32797), Scott D. Stimpson (Reg. No. 33607), William S. Galliani (Reg. No. 33885), Gary S. Williams (Reg. No. 31066), Mark A. Farley (Reg. No. 33170) and Ann L. Gisolfi (Reg. No. 31956), all of Pennie & Edmonds LLP, whose addresses are 1155 Avenue of the Americas, New York, New York 10036, 1667 K Street N.W., Washington, DC 20006 and 3300 Hillview Avenue, Palo Alto, CA 94304, and each of them, my attorneys, to prosecute this application, and to transact all business in the Patent and Trademark Office connected therewith.

SEND CORRESPONDENCE TO:	PENNIE & EDMONDS LLP 1155 Avenue of the Americas New York, N.Y. 10036-2711		DIRECT TELEPHONE CALLS TO: PENNIE & EDMONDS LLP DOCKETING (212) 790-9090		
201	FULL NAME OF INVENTOR	LAST NAME COSTA	FIRST NAME MICHAEL	MIDDLE NAME R.	
	RESIDENCE & CITIZENSHIP	CITY SAN FRANCISCO	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES	
	POST OFFICE ADDRESS	STREET 86 TIOGA AVENUE	CITY SAN FRANCISCO	STATE OR COUNTRY CALIFORNIA	ZIP CODE 94134
202	FULL NAME OF INVENTOR	LAST NAME DOBERSTEIN	FIRST NAME STEPHEN	MIDDLE NAME K.	
	RESIDENCE & CITIZENSHIP	CITY SAN FRANCISCO	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES	
	POST OFFICE ADDRESS	STREET 2 TOWNSEND STREET, APT. 3-205	CITY SAN FRANCISCO	STATE OR COUNTRY CALIFORNIA	ZIP CODE 94107
203	FULL NAME OF INVENTOR	LAST NAME ELSON	FIRST NAME SARAH	MIDDLE NAME L.	
	RESIDENCE & CITIZENSHIP	CITY SAN FRANCISCO	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES	
	POST OFFICE ADDRESS	STREET 1601 FLORIDA STREET	CITY SAN FRANCISCO	STATE OR COUNTRY CALIFORNIA	ZIP CODE 94110
204	FULL NAME OF INVENTOR	LAST NAME FERGUSON	FIRST NAME KIMBERLY	MIDDLE NAME CARR	
	RESIDENCE & CITIZENSHIP	CITY BURLINGAME	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES	
	POST OFFICE ADDRESS	STREET 1038 CHULA VISTA AVE., APT. 4	CITY BURLINGAME	STATE OR COUNTRY CALIFORNIA	ZIP CODE 94010
205	FULL NAME OF INVENTOR	LAST NAME HOMBURGER	FIRST NAME SHEILA	MIDDLE NAME AKIKO	
	RESIDENCE & CITIZENSHIP	CITY OAKLAND	STATE OR FOREIGN COUNTRY CALIFORNIA	COUNTRY OF CITIZENSHIP UNITED STATES	
	POST OFFICE ADDRESS	STREET 69 GLEN AVENUE, APT. 304	CITY OAKLAND	STATE OR COUNTRY CALIFORNIA	ZIP CODE 94611

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

SIGNATURE OF INVENTOR 201 MICHAEL R. COSTA	SIGNATURE OF INVENTOR 202 STEPHEN K. DOBERSTEIN	SIGNATURE OF INVENTOR 203 SARAH L. ELSON
DATE	DATE	DATE
SIGNATURE OF INVENTOR 204 KIMBERLY CARR FERGUSON	SIGNATURE OF INVENTOR 205 SHEILA AKIKO HOMBURGER	
DATE	DATE	